

Spring 2023

## **CS 634: Data Mining**

Mahsa Asadi

Follow this and additional works at: <https://digitalcommons.njit.edu/cs-syllabi>

---

### **Recommended Citation**

Asadi, Mahsa, "CS 634: Data Mining" (2023). *Computer Science Syllabi*. 348.  
<https://digitalcommons.njit.edu/cs-syllabi/348>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Computer Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact [digitalcommons@njit.edu](mailto:digitalcommons@njit.edu).



## **CS 634 - Data mining, Spring 2023**

**Instructor:** Mahsa Asadi

**Office:** GITC 4416

**Email:** [ma2266@njit.edu](mailto:ma2266@njit.edu)

**Sections 002:**

**Class location and time:** Cullimore Hall LECT 2, Mondays 1:00 pm - 3:50 pm

**Sections 004:**

**Class location and time:** Kupfrian Hall 204, Thursdays 10:00 am – 12:50 pm

**Overview:**

This course covers the principles of data mining system design and implementation. It presents methods for association and dependency analysis as well as classification, prediction, and clustering. Optional topics may include time series and graph mining, current trends in data mining, and data mining for scientific, medical and engineering applications.

Data mining is a powerful tool used to discover patterns and relationships in data. We learn how to apply data mining principles to the dissection of large complex data sets, including those in very large databases or through web mining. Explore, analyze and leverage data and turn it into valuable, actionable information. After learning how to analyze data statistically, students learn how to sort through large datasets to identify trends, patterns, and relationships and discover insights previously unknown and to leverage them in business operations.

**Topics Include:**

- Decision trees
- Association rules
- Clustering
- Case-based methods
- Data visualization

**Textbook/ Course Materials:**

The reference book for this course is:

**DATA MINING** Concepts, Models, Methods, and Algorithms  
Mehmed Kantardzic

**Course Policies:**

Attendance is mandatory. A student who misses more than 5 classes will be dropped, without credit.

**Course Communication:**

Canvas (canvas.njit.edu) will be used to post lecture notes, to submit assignments and for course discussion and grades.

**Assignments & Lab:**

All assignments must be submitted via Canvas on the due date. All submitted

work (including exams) must include your name and student ID. It will not be accepted late except for special circumstances (such as jury duty or medical problem), for which you must provide documentation approved by Dean of Students.

Weekly announcements on Canvas will include details for each assignment.

Plagiarism will result in zero credit for the assignment and/or an XF grade in the course.

Cell phones must be turned off during class.

Students will be informed of any modifications of the syllabus during the semester.

### **Material covered:**

- 1) INTRODUCTION: Data-Mining Concepts
- 2) Preparing the data: How to treat the missing values?
- 3) Data Reduction: Dimensionality of the data
- 4) Learning from Data: Classic Machine Learning Algorithms
- 5) Statistical Methods: Regressions and Bayesian Rule
- 6) Decision Trees and Decision Rules: C4.5 Algorithm
- 7) Artificial Neural Networks: Deep Learning (Presentations)
- 8) Cluster Analysis
- 9) Association Rules: Algorithm Apriori
- 10) Web Mining and Text Mining: Recommender Systems (Projects)
- 11) Advances in Data Mining
- 12) Visualization Methods (Final Project)
- 13) Final Exam

### **Evaluation:**

The evaluation will be based on the following course requirements:

- Hands-on/Lab 10%
- Presentation 15%
- Final Project 25%
- Final Exam 50%

The letter grade is based on the overall course score:

Grade formula						
Grade	A	B+	B	C+	C	D
Overall Course Score	89-100	79-88	69-78	59-68	49-58	40-48

## Exam Policies

Final exam TBD.

You must bring a photo ID to all exams. Students with special needs are advised to plan with OARS.

There are no makeup exams. If you miss an exam because of a documented special circumstance you may receive an imputed grade based on the other exam.

If you believe that you deserve more credit than you have been awarded on a particular exam problem, you may request, within 48 hours of the exam being returned, that it be regraded. Your entire exam will be regraded, which may result in points being added or subtracted.

Exams require to bring your own computer. Any other electronic devices, such as cell phones, smart watches, or calculators must be put away and turned off during the exam.

## University Code on Academic Integrity

*“Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at: <http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf>.*

*Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at [dos@njit.edu](mailto:dos@njit.edu)”*

## Tentative Weekly Coverage of Material

The following table shows approximately how much time may be devoted to each topic and the corresponding reading from the suggested textbook.

Week	Topic	Textbook reading	Due work
1	INTRODUCTION: Data-Mining Concepts	Chapter 1	Assignment 1
2	Preparing the data: How to treat the missing values?	Chapter 2	Assignment 2
3	Data Reduction: Dimensionality of the data	Chapter 3	Assignment 3
4	Learning from Data: Classic Machine Learning Algorithms	Chapter 4	Assignment 4
5	Statistical Methods: Regressions and Bayesian Rule	Chapter 5	Assignment 5
6	Decision Trees and Decision Rules: C4.5 Algorithm	Chapter 6	Assignment 6
7	Artificial Neural Networks: Deep Learning (Presentations)	Chapter 7	
8	Cluster Analysis	Chapter 9	Assignment 7
9	Association Rules: Algorithm Apriori	Chapter 10	Assignment 8
10	Web Mining and Text Mining: Recommender Systems (Projects)	Chapter 11	Assignment 9
11	Advances in Data Mining	Chapter 12	Assignment 10
12	Visualization Methods (Final Project)	Chapter 15	
13	Final Exam		