Data Mining/CS 634 Syllabus

Class meetings: According to NJIT course schedule and room assignment.

Instructor Information

Instructor	Email	Office Location & Hours
Dr. Yasser Abduallah	ya54@njit.edu	By appointment via Webex

*I will respond to all emails/Inbox messages within 24 hours. Assignments will be graded, and grades will be returned to students within 2 weeks of their due date, but they may take longer.

General Information

Course Description

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.

Prerequisites/Co-requisites

Fluency in a programming language (Python) is required. Substantial coding and programming will be required in this course.
If you don't have this knowledge and are not willing to learn Python programming language by yourself, you may want to drop the course, but note that learning Python is not difficult. I suggest that you consult with me first before your final decision.

Course Learning Outcomes

By the end of the course, students will be able to:

1. Explain data mining concepts, principles, and methods,
2. Use a wide range of publicly available data mining tools,

3. Evaluate the effectiveness and efficiency of these data mining tools based on different performance measures,
4. Design and implement data mining algorithms, heuristics, methods, and techniques in the context of custom datasets to build custom data mining models and tools.

Course Materials

- Data Mining: Concepts and Techniques, Han et al., Elsevier, 2022, ISBN 9780128117606
- Introduction to Data Mining, Tan et al., Pearson, 2019, ISBN-13: 978-0-13-312890-1.
- Recommended: Murach's Python for Data Analysis, Scott McCoy, Murach Books, ISBN 978-1943872763

Grading Policy

NJIT Grading Legend

Final Grade Calculation

Final grades for all assignments will be based on the following percentages:

Class Participation (Discussion Forums)	**10%**
Knowledge Checks	**5%**
Midterm Project	**25%**
Final Term Project	**20%**
Presentation	**10%**
Final Exam	**30%**

Course Work

Class Participation: (10% of grade) You are expected to participate in weekly (or whenever the assignment is posted and required) discussion forums in Canvas. When all students participate in a discussion, it creates an active learning environment that will help you better understand the materials and be more successful in the class. You will post your initial response to the prompt and respond to two classmates by Sunday at 11:59pm of the week they are listed. In addition, there will be 2 sample quizzes. These sample quizzes will not be graded, but they are designed to familiarize all students with Respondus LockDown Browser and they are part of class participation. The sample quizzes will be proctored as per the policy listed in the section below.

Knowledge Checks: (5% of grade) Knowledge checks questions are designed to assess the performance of the students and elicit performance from the students. You are expected to answer the knowledge checks questions related to the lectures given in each week (or whenever the assignment is posted and required) to better understand the related course modules and be more successful in the class.

Midterm Project: (25% of grade) The midterm project requires you to implement the Apriori algorithm for association rule mining and compare it with a brute force method. The Apriori algorithm and brute force method must be implemented from scratch. You are not allowed to use existing libraries or packages to do the implementation.

Final Term Project: (20% of grade) The final project requires you to implement classification, clustering, or text mining algorithms, which are explained in detail in the course modules.
NOTE: The Projects' requirements may change, if so, the changes and modifications will be posted to the project's submission page.

Presentation: (10% of grade) Each student is required to read a state-of-the-art article on data mining and write a summary, design an example, and describe your thoughts about the article in a PowerPoint presentation. Submit the file as your presentation.

Final Exam: (**30% of grade, mostly open book**) The final exam will contain algorithmic questions, which will require substantial calculations. Therefore, make sure to bring a high-performance calculator to the exam. The final exam will be proctored as per the policy listed in the section below (in case the final exam is online).

Feedback

TA and I will deliver feedback on each assignment via email or using the comments feature in Canvas.

Letter to Number Grade Conversions

A	93-100
B+	86-92
B	78-85
C+	70-77
C	60-69
F	0-59

Exam Information and Policies

Policy for Make-Up Exam: There will be no make-up final exam, except in rare situations where the student has a legitimate reason for missing the exam, including illness, death in the family, accident, requirement to appear in court, etc. The reason must be justified and supported by appropriate documentation. If the student has a conflict with the listed exam date/time, the student must inform me one week before the exam date so that an appropriate arrangement can be made in advance.

Proctoring Policy: In this course you will be required to use the following proctoring method to ensure academic integrity for the exam. See below for more information about how the exam will be proctored in this course. For more information on NJIT's Proctoring Policy, go to the Online Course Exam Proctoring page.

Respondus LockDown Browser and Monitor

Respondus LockDown Browser is a locked browser for taking assessments or quizzes in Canvas. It prevents students from printing, copying, going to another URL, or accessing other applications during a quiz. If a Canvas quiz requires that LockDown Browser be used, students will not be able to take the assessment or quiz with a standard web browser. Students may be required to use LockDown Browser with a webcam (Respondus Monitor), which will record students during an online exam.

The webcam can be built into your computer or can be the type that plugs in with a USB cable. Watch this short video (Links to an external site) to get a basic understanding of LockDown Browser and the webcam feature. A student Quick Start Guide (PDF) (Links to an external site) is also available.

Respondus Lockdown Browser and Monitor does not work with Linux and Chromebooks at this time. Please visit the Respondus Knowledge Base article on computer requirements (Links to an external site) for additional information.

There are two quizzing engines currently available in Canvas, "classic" quizzes and "new" quizzes. Respondus works with both. In CS 634, we use the "classic" quizzes.

Questions or problems can be submitted via web form by going to: **servicedesk.njit.edu** (Links to an external site) and clicking on the "Report your issue online" link. You may also call the IST Service Desk with any questions at 973-596-2900.

More information and Tips for Ensuring a Smooth Experience

Using LockDown Browser with "New" Quizzes in Canvas

When accessing a New Quiz in Canvas using LockDown Browser, students must:

Download and install the LockDown Browser link (provided by the instructor).

Log into Canvas using your standard browser.

Click on the course within your “Courses” list in which you have to take the exam that requires LockDown Browser.

After you enter the course, find the exam and click on it.

A new tab will open with a message stating “Assessment Loading”. You will also see a pop-up window asking you to open Lockdown Browser. Click “Open Lockdown Browser”.

Lockdown Browser will automatically launch and your quiz will be loaded into Lockdown Browser. Click “Begin” to take the quiz. Once a quiz has been started with LockDown Browser, you cannot exit until the “Submit Quiz” button is clicked.

If you are required to use a webcam (Respondus Monitor), you will be prompted to complete a Webcam Check and other Startup Sequence steps.

Once the exam has been submitted, Respondus will close to allow you to continue using your regular browser window.

Policy for Late Work

An assignment is late if it is not submitted to Canvas before the deadline. If you turn in your assignment *n* days late, your total point will be deducted by (50 × *n*) points. For example, suppose you turn in your assignment 1 day late (if you turn in your assignment after the deadline on the due date, it is also considered as 1 day late). Then, you lose (50 × 1) = 50 points automatically, and your total point is 50 points. Further, suppose you lose 10 points in documentation. Thus, you receive (50 –10) = 40 points in total. For all late submissions of the assignment, they must be emailed to me at ya54@njit.edu.

Note: Each student should submit one copy of the assignment only. If the student has submitted his/her assignment (even incomplete) in Canvas, the student is NOT allowed to send another copy of the assignment to ya54@njit.edu . Your assignment will automatically lose 80 points if this rule is violated.

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 NJIT academic code of integrity policy.

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.

Netiquette

Throughout this course, you are expected to be courteous and respectful to classmates by being polite, active participants. You should respond to discussion forum assignments in a timely manner so that your classmates have adequate time to respond to your posts. Please respect opinions, even those that differ from your own, and avoid using profanity or offensive language.

Weekly Expectations

This course is organized by weekly modules in which students will complete multiple modules per week (please see schedule below). Each week, students must watch lecture videos, complete reading assignments (slides files in pdf format) and participate in a class discussion forum by Sunday at 11:59pm.

Course Schedule

#	Topic	Reading	Assignment	Due Dates
1 Module 1, Module 2A, Module 2B	Introduction to Data Mining, Association Rule Mining	Book: *Data Mining: Concepts and Techniques*, Chapters 1, 2, 3, 6 Book: *Introduction to Data Mining*, Chapters 1, 2, 5.1-5.5	Module 1 Knowledge Check Module 2 Knowledge Check Module 1 Discussion Question Sample Quiz	TBA
2 Module 3A, Module 3B, Module 4A, Module 4B	Decision Trees, Other Classification Methods	Book: *Data Mining: Concepts and Techniques*, Chapters 8, 9 Book: *Introduction to Data Mining*, Chapters 3,4	Module 3 Knowledge Check Module 3 Discussion Question Sample Quiz 2 Module 4 Knowledge Check	TBA
3 Module 5A, Module 5B, Module 6A, Module 6B, Module 6C	Naive Bayes, Random Forests	Book: *Data Mining: Concepts and Techniques*, Chapters 8, 8.6 Book: *Introduction to Data Mining*, Chapters 4.4, 4.10	Module 5 Knowledge Check Module 5 Discussion Question Module 6 Knowledge Check	TBA
4 Module 7, Module 8A, Module 8B	Simple Linear Regression, Evaluating Classifier Performance	Book: *Data Mining: Concepts and Techniques*, Chapters 3.4.5, 8.5 Book: *Introduction to Data Mining*,	Module 7 Knowledge Check Module 7 Discussion Question	TBA

		Chapter 4.6	Module 8 Knowledge Check	
5 Module 9A, Module 9B, Module 10A, Module 10B, Module 11A, Module 11B	K-Means Clustering Method, Agglomerative Hierarchical Clustering Method, Cluster Evaluation	Book: *Data Mining: Concepts and Techniques*, Chapters 10, 11, 12, 10.6 Book: *Introduction to Data Mining*, Chapters 7	Module 9 Knowledge Check Module 9 Discussion Question Module 10 Knowledge Check Module 11 Knowledge Check	TBA
6 Module 12	Graph Clustering	Book: *Data Mining: Concepts and Techniques*, Chapters 11.3, 13 Book: *Introduction to Data Mining*, Chapter 8.4	Module 12 Knowledge Check Module 12 Discussion Question	TBA
7 Module 13A, Module 13B	Text Mining	Book: *Data Mining: Concepts and Techniques*, Chapters 11.3, 13	Module 13 Knowledge Check Module 13 Discussion Question	TBA
8 Module 14A, Module 14B, Module 15	Keyword Based Search Engines, Query Based Web Search Systems	Book: *Data Mining: Concepts and Techniques*, Chapter 1.6.2	Module 14 Knowledge Check Module 14 Discussion Question Module 15 Knowledge Check	TBA
9 Module 16A, Module 16B	Data Warehouse and Mediator	Book: *Data Mining: Concepts and Techniques*, Chapter 4	Module 16 Knowledge Check Module 16 Discussion Question	TBA
10 Module 17A, Module 17B	Web Usage Mining	Book: *Data Mining: Concepts and Techniques*, Chapter 10.5	Module 17 Knowledge Check Module 17 Discussion Question	TBA

11 Module 18A, Module 18B, Module 18C	Web Structure Mining	Book: *Data Mining: Concepts and Techniques*, Chapter 10.5	Module 18 Knowledge Check Module 18 Discussion Question	TBA
12 Module 19A, Module 19B	Web Crawling	Book: *Data Mining: Concepts and Techniques*, Chapter 10.5	Module 19 Knowledge Check Module 19 Discussion Question	TBA
13 Module 20A, Module 20B	Time Series Data Mining	Book: *Data Mining: Concepts and Techniques*, Chapter 13.1	Module 20 Knowledge Check Module 20 Discussion Question	TBA
14 Module 21A, Module 21B, Module 21C	Advanced Data Mining I – Association Analysis with FP Tree	Book: *Data Mining: Concepts and Techniques*, Chapter 6.2.4 Book: *Introduction to Data Mining*, Chapter 5.6	Module 21 Knowledge Check Module 21 Discussion Question	TBA
15 Module 22A, Module 22B, Module 22C	Advanced Data Mining II – Scalable Decision Tree Induction	Book: *Data Mining: Concepts and Techniques*, Chapter 8 Book: *Introduction to Data Mining*, Chapter 3.3	Module 22 Knowledge Check Module 22 Discussion Question	TBA

Additional Information and Resources

Canvas Accessibility Statement

Requesting Accommodations:

If you need accommodations due to a disability please contact Scott Janz, Associate Director of the Office of Accessibility Resources and Services, Kupfrian Hall 201 to discuss your specific needs. A Letter of Accommodation Eligibility from the office authorizing student accommodations is required.

General Resources for NJIT Students (including technical support)