Course Syllabus

Jump to Today　　✎ Edit

CS 634 Data Mining Syllabus, Spring 2025

Instructor: Senjuti Basu Roy

Location: KUPF 108

Class sessions: Monday, 1 pm – 3:55 pm

Instructor Office: GITC 4302
Office Hours: Wednesday/Friday (11:45 – 12:45 pm) /by appointment (**email: senjutib@njit.edu (mailto:senjutib@njit.edu)**)

TA: Naidan Zhang

Email of the TA: **nz85@njit.edu (mailto:nz85@njit.edu)** , office hour : Wednesday 3:15-4:15 pm/ Friday 4-5 pm /by appointment

location: 4th floor of GITC (PhD students lab), at workstation 35

Class canvas website: **https://njit.instructure.com/courses/46926 (https://njit.instructure.com/ courses/46926)**

Course Overview: Welcome. The data mining course presents methods and systems for mining varied data and discovering knowledge from data. After detailing a data mining system architecture and tasks, the course examines and compares specific methods in data mining, such as data preparation, classification, clustering, and text mining. Several applications are detailed, and tools to build new applications are provided. The task of knowledge discovery is then outlined as a higher-level goal of data mining. Familiarity with statistics, and database systems; in particular database design is expected.

The **primary objectives** of the course are:

- Understand algorithms and methods of data mining.
- Develop data mining programs and applications.
- Program using available data mining tools and general-purpose languages.
- Understand analysis, metrics, visualization and navigation of data mining results.
- Learn how to use a few commercial data mining tools.

The **outcomes** of the course are:

Upon successful completion of the course, students are able to:

- Explain the basic principles of the primary data mining techniques.
- Explain the difference between data mining, data warehousing, machine learning, etc.
- Design mining models and manage databases to enable data mining technologies as part of larger systems.
- Describe issues facing latest trends in data mining.

Prerequisite: You should be proficient in database design and have an understanding of basic database system implementation techniques. In addition to that, basic understanding in probability and statistics is desirables.

We also recommend that all students have prior experience with at least one programming language such as C/C++/Python and have demonstrated ability to work on coding algorithms and data structures. Students with no programming experience or interest should not take this course.

Texts: Data Mining: Concepts and Techniques (3rd edition), by Jiawei Han, Micheline Kamber and Jian Pei, Morgan Kaufmann, 2011

Additional Readings: The links to additional readings will be made available on canvas during the term. You may bring your readings to each class session so that you can refer to them during discussion. You will additionally do a number of readings of your own choosing for the project and the accompanying research paper that you will write. Your instructor may assign specific readings to individual students.

Course deliverables/grading policies:

Assignment name	Grade
3 homework, each homework will have a theory and a hands on component	10% each, 30% overall (posting dates: Feb 21, March 28, April 25)
Midterm exam	15% March 14

Final Exam	30% TBD
Final project+ presentation	20% (15% +5 %) TBD
Class participation	5%

Homework will always be due at midnight. Submissions more than 6 hours late will be penalized 20% and no submissions will be accepted 24 hours after the due date. **All homework will be distributed and submitted through canvas. You will have 2 weeks to work on each homework. No late submission will be accepted for the final project**

Midterms & final exams: All exams are open book and open notes

Final project: The primary objective of the project is to design an end-to-end data mining solution to a real life problem. The students would be provided with a real-world large dataset and a problem statement. They are required to design the solution, implement it, and perform validation. The solution strategy would involve implementing one or more of the data mining algorithms studied in the class. The students are required to submit the code after appropriate documentation, respective design document, and the result files. They will also have to submit an intermediate project report, as well as prepare a final presentation.

Extra credits will be provided to home works and the exams

*I reserve the right to make small adjustments to grade weights, or to add small assignments as the need arises.

Class Schedule: (may change)

	topics covered	*Readings* *(text; other)*
Week 1		
	Introduction	*Chapter 1*

Week 2	*Getting to know your data*	*Ch.2*
Week 3	*Data Preprocessing*	*Ch. 3*
Week 4	*Data Warehousing and OLAP*	*Ch. 4 & 5*
Week 5	*Frequent Pattern Mining*	*Ch. 6*
Week 7	*Frequent Pattern Mining*	*Ch. 6*
Week 8	*Classification basics*	*Ch. 8*
Week 9	*Clustering*	*Ch. 10*
Week 10	*Clustering*	*Ch. 10 & 11*
Week 11	*Outlier Detection*	*Chapter 12*
Week 12	*Emerging topics*	*Chapter 13*
Week 13	*Graph mining and social network analysis*	*Reading material will be provided*

| Week 14 | *Web search and Information Retrieval* | *Reading material will be provided* |
| Week 15 | *Student Presentation* | *Both days* |

Grading Scale

A: 93% and above; B+: 86%-92.9%; B: 78%-85.9%; C+: 70%-77.9%; C: 60%-69.9%; F: Below 60%.

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: **NJIT Academic Integrity Code ⬚ (https://t.e2ma.net/click/rh9gwmb/v0mfoovf/v02nt2x)** *.*

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)** *."*

Use of Generative AI: Complete Prohibition : *This course expects students to work without artificial intelligence (AI) assistance in order to better develop their skills in this content area. As such, AI usage is not permitted throughout this course under any circumstance.*

Student Absences for Religious Observance: NJIT is committed to supporting students observing religious holidays. Students must notify their instructors in writing of any conflicts between course requirements and religious observances, ideally by the end of the second week of classes and no later than two weeks before the anticipated absence. Alternative accommodations would be provided for missed assignments, exams, quizzes, or other coursework within the term.

Course Summary:

Date	Details	Due

Date	Details	Due
Tue Jan 28, 2025	📝 **Academic Engagement: Spring 2025 (https:// njit.instructure.com/courses/46926/ assignments/512786)**	due by 11:59pm