New Jersey Institute of Technology
Ying Wu College of Computing
Computer Science Department
Data Mining

Code: CS634
Mode: Face-to-Face
Instructor: Khalid Bakhshaliyev
Office: TBA
Email: khalid.bakhshaliyev+cs634@njit.edu
Office Hours: TBD
Note: Your messages will be answered by the end of the next day. Grades for all items will be getting posted during the week after their due date. For issues with your grades, contact the grader and cc the instructor.
Teaching Assistant: Rudranil Maity - rm964@njit.edu

Tutoring. NJIT provides a tutoring service. Please contact one of the available tutors. Their availability and contact information can be found here.

Course Description

Data Mining is a fundamental discipline in the field of data science and analytics, focused on extracting valuable insights and patterns from large datasets. This course provides students with a comprehensive introduction to the principles, techniques, and applications of data mining. Through a combination of theoretical lectures, hands-on exercises, and real-world case studies, students will gain the knowledge and skills necessary to uncover hidden knowledge within data.

Prerequisites

The course does not have other course prerequisites.

Background on some basic calculus, linear algebra, probability and programming ability is required. The following free online materials are recommended for reviewing this background:
Mathematics for Machine Learning
A visual guide to NumPy

Course Textbooks

"Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. (https://www.pearson.com/en-us/subject-catalog/p/introduction-to-data-mining/P200000003204/9780137506286)
"Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei.
“Mining Massive Datasets” by Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman

---

Coursework, Assessment and Related Outcomes

Hands-On Assignments [25%]. Six hands-on assignments of equal grading weight. The weakest of the six grades is dropped from the calculation.
Quizzes [15%]. Weekly quizzes reinforcing the material of each module.
Project midterm presentation [15%]. Presentation of the collected data, preliminary work and hypothesis. Grading will be based on how interesting the data is, good layout of the preliminary work done by others, what ways are used (or intended to be used) to collect the data, scope of the project and commitment to the duration allocated for each presentation.
Project final presentation [30%]. Presentation of the proof or disproof of the hypothesis; the methods used to preprocess the data; algorithms used to derive the knowledge from the data. The grades will be according to the complexity of the project, visualization, knowledge we learn from the data, and commitment to the duration for the allocated time for each presentation.
Project report[15%]**.** Report should include Introduction, Preliminary work, Methodology, the knowledge (Proof or disproof of the hypothesis), and conclusion. The project should be reproducible, the link to the data and the code (github or any other repository) should be referenced in the report.

Project: Groups up to 2 students allowed.

Feedbacks on presentations[Extra 7%]. Students' feedback will be collected on each presentation (How the timing was, how interesting the presentation is, how knowledgeable the presenter is about the topic, the knowledge the presentation provides etc.)

Participation in the weekly discussions[5%]. Students are encouraged to discuss weekly topics; and also discuss the open data sources they are intending to use to give each other ideas.

Grading Scheme and Letter Grades.
A: [93-100], B+: [86-93), B:[78,86), C+: [70,78), C: [60,70), F:<60

Grading Feedback. Assignment marks will be accompanied with solutions and general feedback summarizing common mistakes. Individual grading feedback will be given whenever possible. Further clarifications can be provided via direct communication with the instructor and the course grader.

Late Work Policy. In the case when a student is unable to complete an assignment or other serious reasons, these must be communicated and documented promptly. In any other case, each hour of delay after the due date will incur a 2% score reduction. No extensions will be granted. However, the lowest assignment score and the two lowest quiz assignments scores for each student will be dropped.

Course Topics

1. **Data**
2. **Classification - Basic concepts and techniques, Alternative Techniques**
3. **Association Analysis - Basic Concepts and Algorithms, Advanced Concepts**
4. **Cluster Analysis - Basic Concepts and Algorithms, Advanced Issues**
5. **Anomaly Detection**
6. **Avoiding False Discoveries**
7. **Mining Data Streams**
8. **Graph Mining Techniques - PageRank**
9. **Mining Social-Network Graphs**
10. **Unsupervised Deep Learning**

Course Policies

Email

Use of your NJIT email or Canvas inbox is strongly encouraged.

Grade Corrections

Check the grades in course work and report errors promptly. Please try and resolve any issue within one week of the grade notification.

Exam and Proctoring Policy

See the NJIT Online Course Exam Proctoring page for information on proctoring tools and requirements.

Incomplete

A grade of I (incomplete) is given in rare cases where work cannot be completed during the semester due to documented long-term illness or unexpected absence for other serious reasons. A student needs to be in good standing (i.e. passing the course before the absence) and receives a provisional **I** if there is no time to make up for the documented lost time; an email with a timeline of what is needed to be done will be sent to the student. Note that an **I** must always be resolved by the end of the next semester.

Collaboration and External Resources for Assignments

Some homework problems will be challenging. You are advised to first try and solve all the problems **on your own**. For problems that persist you are welcome to talk to the course assistant or the instructor. You are also allowed to collaborate with your classmates and search for solutions online. But you should use such solutions only if you understand them completely (admitting that you don't understand something is way better than copying things you don't understand). Also make sure to give the appropriate credit and citation.

Requesting Accommodations

If you need an accommodation 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.

NJIT Services for Students, Including Technical Support
Please follow this link.

Canvas Accessibility Statement
Please follow this link.

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 a this link.
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