CS 634 Data Mining

Instructor
Name: Dr. Jason Wang
Office: GITC 4211
Phone: (973) 596-3396
Email: wangj@njit.edu

Test Proctoring Requirements
The final exam will be conducted in the Respondus LockDown Browser+Webcam environment. Visit CS 634 Canvas and https://ist.njit.edu/respondus/ to get details of Respondus.

There will be 2 sample quizzes that help students understand the proctoring, style and format of the final exam so as to better prepare for the exam.

Prerequisite and Required Background
Fluency in a programming language (Python or Java) is required. Substantial coding and programming will be required in this course.

Course Description (from NJIT's Graduate Catalog)
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.

Course Objectives/Outcomes
To familiarize students with basic data mining principles, data mining methods and tools, as well as advanced data mining

applications, and to help students find jobs in the field of data mining and machine learning. Upon completion of the course, students will be able to

- Explain data mining and machine learning concepts, principles and methods,
- Use a wide range of publicly available data mining and machine learning tools,
- Evaluate the effectiveness and efficiency of these data mining and machine learning tools based on different performance measures,
- Design, develop and implement custom data mining and machine learning algorithms, heuristics, methods, techniques and software tools.

Course Textbooks

- Data Mining: Concepts and Techniques, Han et al., Elsevier, 2011, ISBN 978-0-12-381479-1. [Edition 4, 2022]
- Introduction to Data Mining, Tan et al., Pearson, 2019, ISBN-13: 978-0-13-312890-1.
- Deep Learning, Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press, 2016.

Course Outline and Schedule
This course covers the concepts and techniques of

1. Association Rule Mining

2. Classification, Regression, Prediction and Forecasting (Supervised Machine Learning)

3. Clustering (Unsupervised Machine Learning)

4. Time Series Data Mining

5. Graph Mining

6. Current Trends in Data Mining (Deep Learning)

7. New Applications (Deep Learning Applications)

Course Workload
There will be one midterm project, one final term project, one term paper, and one final exam.

Course Grade
Midterm Project -- 30%, Final Term Project -- 25%, Term Paper -- 10%, Final Exam -- 35%.

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%.

Honor and Policy

- Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an "F" grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified.
- In the exam, each student is required to sign the Honor Code Agreement "On my honor, I pledge that I have not violated the provision of the NJIT Student Honor Code."