New Jersey Institute of Technology
Digital Commons @ NJIT

Computer Science Syllabi | NJIT Syllabi

Spring 2020

CS 634-102: Data Mining

Hseu-Ming Chen

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

Recommended Citation
Chen, Hseu-Ming, "CS 634-102: Data Mining" (2020). *Computer Science Syllabi*. 57.
https://digitalcommons.njit.edu/cs-syllabi/57

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.

CS-634 Data Mining (Spring 2020)

General Information

Location:	CKB 215
Time:	Thursday 6pm - 8:50pm
Instructor:	Dr. Hseu-Ming Chen
Office Hours:	By appointment only - preferably between 3:30pm and 5:30pm on Thursdays Location: GITC 4422 (meeting room across from the seminar room)
Email:	hseuming@njit.edu hseu.chen@njit.edu
Grader:	Sahaj Marwah; ssm226@njit.edu ; Office hours: 5:30pm - 6:30pm on Mondays @ GITC 4403
Class web page:	Download class notes and homeworks from Canvas

Course Overview and Objectives

This course will provide an overview of the applications, methods, tools and technologies that constitute data mining. Due to the diversity of subjects that comprise this field, the class will inevitably have more breadth than depth. At the beginning of the course we will cover 'core' data mining topics, such as basic data mining methods. After this material is covered, the focus of the class will shift to surveying applications of data mining and to class projects, which will be a major component of the course. These projects will allow you to pursue your own interests.

Briefly, in this course we will:

- Introduce students to the basic concepts and techniques of Data Mining
- Present dozens of algorithms and implemenation examples for use in real-world and large-scale data mining projects
- Learn about the methods underlying modern data mining, and how to use them
- Improve and clean data quality for reporting and analytics for decision makers
- Develop skills of using recent data mining software for solving practical problems
- Survey applications of data mining
- Explore the future and implications of data mining/science

Specific topics will include:

- Supervised and unsupervised learning
- Data processing and data mining workflows
- Applications of data mining

Prerequisites

Basic understanding of RDBMS, fundamental algorithms, probability, statistics, and familiarity with some programming language(s). If in doubt about the prerequisites, please consult with the instructor for permission to take the class.

Textbooks

Primary Textbook

- Data Mining: Concepts and Techniques (3rd Edition, 2012)
  Jiawei Han, Micheline Kamber, Jian Pei
  ISBN 978-0-12-381479-1
  Elsevier

Reference Textbooks (optional)

- Introduction to Data Mining (2nd Edition, 2019)
  Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar
  ISBN 978-0-13-312890-1
  Pearson
  -- This is a POD book. For on-line version, please contact vitalsource.com.

- Mining of Massive Datasets (3rd Edition, 2020)
  Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman
  ISBN-13: 978-1-108-47634-8
  Cambridge University Press
  -- http://www.mmds.org/

Class Requirements and Grading

The class grade will be based on the following components:

- Homework, computing lab exercise - 15%
- Term project (due at the beginning of the last class) - 25%
- Closed-book in-class midterm - 25%
- Open-book in-class final - 35%

Course Project

Data mining (or data science) is not possible without some sort of programming knowledge. This class will involve hands-on assignments and demonstrations of executing data mining techniques with some programming language and most likely its associated libraries. Understanding of Python, Java, or another similar programming language will be useful, but is not required. You are also welcome to use any other language or toolkit you are comfortable with, but please talk to me first.

The final project will constitute a large part of your grade. You will work with me to select an appropriate final project: but, broadly, this will involve working with a dataset to perform an analysis or accomplish a specific task of interest (e.g., clustering classic literature based on word counts; building a predictive model in a domain that interests you; etc.). One component of your final project grade will be a presentation of the work that you will give in class.

Put succinctly: the course project is an opportunity for student groups to investigate a data mining problem that interests them. The course project should apply data mining techniques to real-world problems. Data and software for these projects can be obtained from various internet sites, or developed by students.

We will further discuss the project requirements and peer evaluation/assessment in class.

A presentation of each project is required in addition to a written report.

Late Assignment Policy

Written Homework and Programs are due at the beginning of class. 25% penalization per late day; Not accepted more than 3 days late.

Attendance

You are supposed to attend all the classes. Participation is highly encouraged to make the class more interactive. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class be sure to consult one of your classmates about the content of the lecture and visit the course web page to get notes, exercises, assignments, deadlines and announcements.

Collaboration and Honor Code

Each student is responsible for his/her own work. Students may discuss problems together but must write up their own solutions. Collaboration on homeworks and programs should be limited only to answering questions that can be asked and answered without using any written medium (e.g., no pencils, instant messages, or email). When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note in particular that copying homework or programming assignments, in full or in part is forbidden. 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.

Modifications to Syllabus

The syllabus may be modified at the discretion of the instructor or in the event of extenuating circumstances. Students will be notified in class of any changes to the syllabus.