

CSC/IE 7810

Section 001

Data Mining: Algorithms and Applications

Winter 2023

M W 5:30 P.M. – 6:45 P.M.

Faculty contact information:

Name: Dr. Suzan Arslanturk

Office address: 14200.10

Office hours: By appointment

Phone: 313-577-6782

Email: gl9815@wayne.edu

Course Description:

With the advent of various swift data acquisition systems and recent developments in the internet technology, huge amounts of data have been amassed in different forms. The bursting need for identifying some interpretable and valuable information from these large datasets has never been more important than it is today. Data mining deals with the development of robust algorithms for nontrivial extraction of hidden and potentially useful information from massive amounts of data. In the last decade, data mining has emerged as one of the most promising and challenging areas in computer science. This course is mainly designed for beginning/senior graduate students who are interested in understanding and applying various data mining concepts to real-world problems such as Healthcare Informatics, Retail and Consumer Datasets, Social Networks and Web. This course will be very useful to students whose interests are in the areas of database systems, data mining, machine learning, healthcare informatics, pattern recognition, information retrieval, bioinformatics and artificial intelligence.

Credit Hours:

3 Credit Hours (Lect. 3)

Prerequisite:

CSC 5800 with grade of C or better

This is a graduate-level course. Some basic background in data structures, algorithms and programming is assumed. Knowledge about data mining/pattern recognition/machine learning/artificial intelligence is absolutely required.

Co-requisites:

None

Required and optional textbook(s):

There is no single text book that will be followed from this course. The material for this course will be taken from a wide variety of resources ranging from books, tutorials and online lecture slides.

Recommended Books:

- Charu C. Aggarwal, ***Data Mining: The Textbook***, Springer, 2015.
- Jiawei Han and Micheline Kamber, ***Data Mining: Concepts and Techniques***, Second Edition, Morgan Kaufmann, 2006.
- Richard O. Duda, Peter E. Hart, and David G. Stork, ***Pattern Classification***, John Wiley & Sons, 2012.
- Stephan P. Kudyba, ***Healthcare Informatics: Improving Efficiency through Technologies, Analytics and Management***, CRC Press, 2016.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman, ***The Elements of Statistical Learning; Data mining, Inference and Prediction***, 2nd Edition, Springer Verlag, New York, 2009.
- Christopher M. Bishop, ***Pattern Recognition and Machine Learning***, Springer, 2007.
- Kevin P. Murphy, ***Machine learning: a probabilistic perspective***, The MIT Press, 2012.
- Pang-Ning Tan, Michael Steinbach and Vipin Kumar, ***Introduction to Data Mining***, Addison-Wesley, First Edition, 2005.
- Jure Leskovec, Anand Rajaraman, Jeff Ullman, ***Mining of Massive Datasets***, 2nd Edition, Cambridge University Press, 2014.

Computer Programs:

RStudio

RStudio is freely available online at: <https://www.rstudio.com/>

Course contents:

The following topics will be covered in this course:

- Overview of Datasets and Information Systems: Healthcare Informatics, Retail and Consumer Datasets, Social Networks and Web
- Text Mining and Applications in Clinical Information Systems
- Web Data Mining
- Evolutionary Data Mining
- Social Network Analysis
- Recommendation Systems and Clinical Applications
- Visual Data Mining
- Hidden Markov Models
- Semi-Supervised Learning and Applications
- Dimensionality Reduction and Applications (Healthcare, Consumer/Retail)
- Graph Mining and Applications (Healthcare, Consumer/Retail)

Course Learning Objectives:

The course is designed to help students:

CSC 7810 Course Learning Objectives

- 1 Demonstrate a comprehensive understanding of different data mining tasks and the algorithms most appropriate for addressing them.
- 2 Actively use Data Mining algorithms and apply data mining tools to real-world problems.
- 3 Select and implement appropriate data mining methods for various datasets and prediction problems in healthcare, retail/consumer, social networks, and web.
- 4 Summarize and evaluate research and development issues.
- 5 Design and implement new algorithms for specific Data Mining application.
- 6 List and describe the recent trends and open directions in the field of data mining.
- 7 Review and examine different Data Mining tools

Assessment:

There will be no exams in the course. Term-project will be one of the major components of this course. Final grades are based on the performance in homeworks, in-class quizzes, paper presentation, and the final project. Here is the distribution:

- 35% Homework Assignments
- 15% Participation & Quizzes
- 15% Paper Presentation
- 35% Course Project

Homework Assignments:

There will be four written homework assignments. Homework problems might constitute some programming exercises that are designed to understand some of the algorithms explained in the class. Students are encouraged to talk and discuss with other students to improve their conceptual understanding, but the final submission must be their own work.

Quizzes:

Each quiz will consist simple questions based on the notes and lectures' slides. The questions asked in these quizzes will be only high-level concepts and basic ideas and will not include any intensive problem solving.

Paper Presentation:

Students are required to select a term project and prepare a 15 minutes presentation for the class (including Q & A). The purpose of this paper presentation is to help students to practice giving talks in front of public at conferences or other situations, and better prepare the course project.

Final Project:

One of the major components of this course is the course project. Students will choose a topic and investigate some interesting solutions (that can be some theoretical analysis, algorithmic implementation/comparison or building a data mining system) in the general area of data mining applied to that particular problem domain. The main purpose of this project is to enable students to get some hands-on experience in the design and implementation of feasible solutions to computationally challenging problems. More details about the project proposal/presentation and project submission will be provided later on in the course.

Grading Scale:

The grades for the course will be based upon the percentages given below

A	90 - 100%	C	70 - 73%
A-	87 - 89%	C-	67 - 69%
B+	84 - 86%	D+	64 - 66%
B	80 - 83%	D	60 - 63%
B-	77 - 79%	D-	57 - 59%
C+	74 - 76%	F	0 - 56%

Grading Policies:

1. There will be no in-completes given for the course.
2. Cheating of any kind is not allowed and will be handled in accordance with University Policy.
3. Otherwise stated, all assignments will be submitted on Canvas by scheduled due dates. Any submission after the due date will be considered late. Late submissions will be penalized by the reduction of the grade by 25% for each late day. A penalty of 25% of the assignment value per day (including weekends) is assessed on late submissions beginning on the day due. Grades will be posted on Canvas within two weeks after assignment due date.
4. You are responsible to check your grades after each assignment, and report an inconsistent grade to the instructor no later than 7 days after the grade was assigned. After 7 days from posting it on Canvas, the grade will become final.
5. All assignments must be submitted using Canvas. No email or hard copy is accepted. You must follow the following format:
 - a. Use a word file to type your answers.
 - b. Include the following information:
 - ✓ Full name
 - ✓ Class name (CSC 7810)
 - ✓ Assignment number and date
 - c. State your answer clearly.
 - d. For programming assignment, include the source file for each problem; use only the code and the software associated with the required book.
 - e. If your assignment requires more than one file include all files in one folder and compressed (zipped) your folder.
 - f. Submit your file to the Canvas. You cannot submit your file more than one time.

If you do not attach the document before submitting, your instructor will not receive your assignment. It is your responsibility to make sure that your file is uploaded correctly.

6. All students are requested to access their Wayne State e-mail account regularly. You may be contacted when important matters arise. If you have any questions about the course or need assistance, please contact the instructor and/or the TA in person during office hours or by e-mail at any time

Religious Holidays:

Because of the extraordinary variety of religious affiliations of the University student body and staff, the Academic Calendar makes no provisions for religious holidays. However, it is University policy to respect the faith and religious obligations of the individual. Students with classes or examinations that conflict with their religious observances are expected to notify their instructors well in advance so that mutually agreeable alternatives may be worked out.

Student Disabilities Services:

- If you have a documented disability that requires accommodations, you will need to register with Student Disability Services for coordination of your academic accommodations. The Student Disability Services (SDS) office is located in the Adamany Undergraduate Library. The SDS telephone number is 313-577-1851 or 313-202-4216 (Videophone use only). Once your accommodation is in place, someone can meet with you privately to discuss your special needs. Student Disability Services' mission is to assist the university in creating an accessible community where students with disabilities have an equal opportunity to fully participate in their educational experience at Wayne State University.
- Students who are registered with Student Disability Services and who are eligible for alternate testing accommodations such as extended test time and/or a distraction-reduced environment should present the required test permit to the professor at least one week in advance of the exam. Federal law requires that a student registered with SDS is entitled to the reasonable accommodations specified in the student's accommodation letter, which might include allowing the student to take the final exam on a day different than the rest of the class.

Academic Dishonesty - Plagiarism and Cheating:

Academic misbehavior means any activity that tends to compromise the academic integrity of the institution or subvert the education process. All forms of academic misbehavior are prohibited at Wayne State University, as outlined in the Student Code of Conduct (<http://www.doso.wayne.edu/student-conduct-services.html>). Students who commit or assist in committing dishonest acts are subject to downgrading (to a failing grade for the test, paper, or other course-related activity in question, or for the entire course) and/or additional sanctions as described in the Student Code of Conduct.

- **Cheating:** Intentionally using or attempting to use, or intentionally providing or attempting to provide, unauthorized materials, information or assistance in any academic exercise. Examples include: (a) copying from another student's test paper; (b) allowing another student to copy from a test paper; (c) using unauthorized material such as a "cheat sheet" during an exam.
- **Fabrication:** Intentional and unauthorized falsification of any information or citation. Examples include: (a) citation of information not taken from the source indicated; (b) listing sources in a bibliography not used in a research paper.
- **Plagiarism:** To take and use another's words or ideas as one's own. Examples include: (a) failure to use appropriate referencing when using the words or ideas of other persons; (b) altering the language, paraphrasing, omitting, rearranging, or forming new combinations of words in an attempt to make the thoughts of another appear as your own.
- **Other** forms of academic misbehavior include, but are not limited to: (a) unauthorized use of resources, or any attempt to limit another student's access to educational resources, or any attempt to alter equipment so as to lead to an incorrect answer for subsequent users; (b) enlisting the assistance of a substitute in the taking of examinations; (c) violating course rules as defined in the course syllabus or other written information provided to the student;

- (d) selling, buying or stealing all or part of an un-administered test or answers to the test;
- (e) changing or altering a grade on a test or other academic grade records.

Course Drops and Withdrawals:

In the first two weeks of the (full) term, students can drop this class and receive 100% tuition and course fee cancellation. After the end of the second week there is no tuition or fee cancellation. Students who wish to withdraw from the class can initiate a withdrawal request on Pipeline. You will receive a transcript notation of WP (passing), WF (failing), or WN (no graded work) at the time of withdrawal. No withdrawals can be initiated after the end of the tenth week. Students enrolled in the 10th week and beyond will receive a grade. Because withdrawing from courses may have negative academic and financial consequences, students considering course withdrawal should make sure they fully understand all the consequences before taking this step. More information on this can be found at:

<http://reg.wayne.edu/pdf-policies/students.pdf>

Student services:

- The Academic Success Center (1600 Undergraduate Library) assists students with content in select courses and in strengthening study skills. Visit www.success.wayne.edu for schedules and information on study skills workshops, tutoring and supplemental instruction (primarily in 1000 and 2000 level courses).
- The Writing Center is located on the 2nd floor of the Undergraduate Library and provides individual tutoring consultations free of charge. Visit <http://clasweb.clas.wayne.edu/> writing to obtain information on tutors, appointments, and the type of help they can provide.

Class recordings:

Students need prior written permission from the instructor before recording any portion of this class. If permission is granted, the audio and/or video recording is to be used only for the student's personal instructional use. Such recordings are not intended for a wider public audience, such as postings to the internet or sharing with others. Students registered with Student Disabilities Services (SDS) who wish to record class materials must present their specific accommodation to the instructor, who will subsequently comply with the request unless there is some specific reason why s/he cannot, such as discussion of confidential or protected information.