WENTWORTH INSTITUTE OF TECHNOLOGY
College of Engineering and Technology

Data Mining

Summer 2023

Course Number: COMP-5705

Instructor Name: Salem Othman

Classroom: Online

Office Location: Williston Hall - 110

Class Schedule: T (5:00 pm-6:20 pm)

Appointments and Meetings: By Appointment

Lecture/Lab/Total Credits: 3/0/3

Office Telephone Number: 617-989-4508

Email address: othmans1@wit.edu

COURSE DESCRIPTION:

Data Mining is the process of finding hidden patterns and rules in large datasets. This course is a graduate level survey of basic concepts, methods, tools, and techniques related to data mining. Topics include data preprocessing; data warehousing and online analytical processing; data cube technology; mining frequent patterns, associations, and correlations; advanced pattern mining; and outlier detection.

COURSE PREREQUISITES/COREQUISITES:

Prerequisite: COMP5700 Classical Artificial Intelligence

REQUIRED TEXTBOOK(S):

1. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition by Jiawei Han (Author), Micheline Kamber (Author), Jian Pei (Author). ISBN-13: 978-9380931913 ISBN-10: 9780123814791.

THE COLLEGE BOOKSTORE:

Location: 103 Ward Street Boston MA 02115
Telephone: 617-445-8814

RECOMMENDED LEARNING MATERIALS:

- Introduction to Data Mining (2nd Edition). By Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. Pearson. 2019. ISBN-13: 978-0133128901 ISBN-10: 0133128903.
- Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition by Jiawei Han (Author), Micheline Kamber (Author), Jian Pei (Author). ISBN-13: 978-9380931913 ISBN-10: 9780123814791
- Any recently published book that covers the topics described in the weekly schedule may supplement the required texts.

- A graphing tool that you might consider: https://www.desmos.com/calculator

COURSE LEARNING OUTCOMES:

At the end of the course the student will be able to:

1. Identify hidden patterns in datasets.
2. Evaluate different models that can be used for data preprocessing.
3. Categorize and compare situations for applying different Data Mining techniques.
4. Propose Data Mining algorithms for different applications.

INSTRUCTIONAL METHODOLOGIES:

This course will combine traditional lecturing with hands-on assignments that reinforce the lecture material. In particular, lectures will focus on concepts and ideas while the assignments will provide concrete experience and skills. Students will complete a final project that requires in-depth use of concepts and skills learned throughout the semester.

ATTENDANCE POLICY:

Students are expected to attend classes regularly, take tests, and submit papers and other work at the times specified by the instructor. Due to the fast pace and nature of the course, attendance of all class meetings is paramount. Although there are perfectly understandable reasons for missing class, all absences - regardless of how well motivated - have a negative effect on the class as a whole. Consequently, this course enforces a strict attendance policy. Attendance is taken in every course meeting. Students who miss 3 course meetings will be provided a warning by the instructor. Students who miss 5 course meetings will receive a failing grade in the course.

GRADING POLICY:

Assessment and Evaluation

Students will complete in-class exercises in every class. There will be 4 assignments total. The last several lectures of the course will be dedicated to a term project. Students will work on the project in teams of 2-3 students and will be required to present their completed project to the class. They will also be required to submit a written report at the end of the semester. There will be a total of two exams.

In-Class Exercises		5%
Assignments		40%
Exam 1 (In Class)		10%
Exam 2 (In Class)		15%
Final Paper and Project	Proposal & Progress (2%)	30%
	Presentation (5%)	
	Final Report (20%)	
	Poster (3%)	

Project proposal:

The project proposal should consist of a 2-page discussion of how you envision your project. You will need to explain the following in the project proposal:

1. What you plan to implement.
2. Why this project is of interest to you.
3. How you plan to implement your project in stages.

The proposal must also include a tentative, detailed time-line explaining what you will be accomplishing each week. A staged approach is CRITICAL to your success. You need to be able to complete a portion and test it before moving on. Always keep a working copy of your project safe (multiple backups). Use proper scientific style for both your proposal and your final report. Keep in mind that this is just the proposal. If you project starts taking a different turn than what you had originally proposed, that is okay. But keep the instructor informed as things change.

Final Report:
Your final report (paper) should be in scientific style, using as many pages as you need to fully explain your project. It is very unlikely that this will be less than 4 pages, and it could be considerably more. Your report should include the sections of introduction to the problem, relevant work, description of your project, results and evaluation, conclusion, Future Work, and references. It is possible that some papers will add more sections or combine some of the sections above.

Code:
You must implement your project and submit your source code along with the final report. You need to clearly document what your work is and what libraries or tools that others have created are. Your code should be substantially original. If you use some code from any other source, it should be clearly marked as not your own, along with a citation of where it came from. If you have found code that you would like to use as part of your project, the safest course of action would be to discuss it with your instructor before you rely on the code. Make sure your code is also well organized and commented. It should not only work correctly but should be easy to read.

Poster:
Your poster should serve as an advertisement for your project. It should include your name, your instructor's name, the course name, an abstract, and some visual material to make it interesting. The poster should be eye-catching and easily readable from 3-5 feet away. Don't forget to also include the title of the project, your name, and the time of the presentation (if known) on your poster.
A good poster will not have too much text nor contain too little information. I would suggest you have a motivational section to help readers understand the importance of your work, a description section which tells about your project's details of implementation (at a somewhat general level), a results and analysis section where you describe how your result data compares to others work, a related work section, and a citations section.
Very nice posters can be made using the large-format printer. To do this, you need to create a PowerPoint presentation (one slide). Depending on the amount of text you have, sizes from 36 points to 40 points might be appropriate for the poster. Please don't go smaller than 18-point font (for reasons mentioned above).

Final Project:

The final project will be done in groups of two or three members. The rules follow:

1. It is your responsibility to find and work with your teammates.
2. The project report will be due at the end of the course. Each group will also be giving a 10-minute presentation to the class at a time selected or allocated to you.

The purpose of the course project is for you to explore how data mining techniques learned in class can be applied to real life problems. The topic of the project is flexible, as long as it is related to data mining.

WENTWORTH GRADING SYSTEM:

Grade	Weight	Numerical Definition	Definition
A	4.00	93-100	Distinction
A-	3.67	90-92	High Pass
B+	3.33	87-89	Pass
B	3.00	83-86	Pass
B-	2.67	80-82	Provisional
C+	2.33	77-79	Provisional
C	2.00	73-76	Provisional
F	0.00	0-72	No Pass
W	0.00		Withdrew

Grades of "P" or "F" are awarded to courses with this grading scheme and carry academic credit. "P" or "F" grades do not calculate into the GPA.

Wentworth does not offer students the option to audit a course; if a student is granted an exception to this policy the course cannot be converted at any time to a credit-bearing course and will not satisfy a degree requirement.

ADD/DROP:

Students should check the academic calendar to confirm the add/drop deadline. Dropping and/or adding courses is done online. Courses dropped in this period are removed from the student's record.

Non-attendance does not constitute dropping a course. If a student has registered for a course and subsequently withdraws or receives a failing grade in its prerequisite, **then the student must drop that course**. In some cases, the student will be dropped from that course by the Registrar. However, it is the student's responsibility to make sure that he or she meets the course prerequisites and to drop a course if the student has not successfully completed the prerequisite. The student must see his or her academic advisor or academic department chair for schedule revision and to discuss the impact of the failed or withdrawn course on the student's degree status.

MAKE-UP POLICY:

All assignments have a specific due date and time. **Assignments that are submitted late will not be accepted.**

Students who miss scheduled exams will not, as a matter of course, be able to make up those exams. If there is a legitimate reason why a student will not be able to complete an assignment on time or not be present for an exam, then they should contact the instructor beforehand. Under extreme circumstances, as decided on a case-by-case

basis by the instructor, students may be allowed to make up assignments or exams without first informing the instructor.

ACADEMIC SUPPORT:

The Center for Academic Excellence facilitates Wentworth students' academic success and helps them to achieve their full learning potential. Students may choose to receive individual assistance through one-on-one tutoring in many subjects, including math, science, writing, and major classes. In addition, the Center for Academic Excellence offers Facilitated Study Groups (FSGs), tutor-led study tables, academic workshops, and learning-strategy consultations. The peer-tutoring program is certified by the College Reading and Learning Association's International Tutor Training Certification program. To make an appointment or to review our drop-in offerings, please visit www.wit.edu/cae. For additional assistance or support on subjects not listed, please reach out via email at cae@wit.edu.

ACADEMIC HONESTY STATEMENT:

Students at Wentworth are expected to be honest and forthright in their academic endeavors. Academic dishonesty includes but is not limited to cheating, prohibited collaboration, coercion, inventing false information or citations, plagiarism, tampering with computers, destroying other people's coursework or lab or studio property, theft of course materials, posting coursework/course materials to websites, or other academic misconduct. If you have any questions, contact your professor prior to submitting an assignment for evaluation. See your academic catalogue for a full list of definitions and the WIT Academic Honesty website for the procedures: wit.edu/academic-honesty.

STUDENT ACCOUNTABILITY STATEMENT:

Any attempt to pass off another's work as one's own is plagiarism.

In this course the penalty for plagiarism is a failing grade in the course for any parties concerned. It is permissible for students to discuss the nature of an assignment or how to use a particular feature of the software. However, not a single keystroke of the work you submit should be done by anyone but you, nor should your work be based on commands supplied by someone else or developed in collaboration with someone else. In other words, you should not sit down and work together with anyone else on the assignments. Nor should you give, receive, or solicit specific information (such as code, commands) from other students in this course. (This, of course, does not apply to assignments that are explicitly assigned to a group.) Exchange of detailed information about an assignment is cheating and will not be tolerated. If you are unsure or unclear about the rules, please contact me.

THE CENTER FOR WELLNESS:

College can be challenging and it is common to feel overwhelmed or stressed at times. If these feelings are related to course work or academic performance, please talk to me. For more significant mental health concerns, **The Center for Wellness (003 Watson Hall, 617-989-4390)** provides free and confidential mental health counseling.

If you or someone you know needs support around thoughts of suicide, the following resources are available:

- The Center for Wellness, Watson 003, 617-989-4390, M-F 8:15-4:45
- Campus Police, First level of 610 Huntington Avenue, 617-989-4444, 24/7
- Samaritans, call or text 1-877-870-4673
- Crisis Text Line, text "start" to 741-741
- National Suicide Prevention Lifeline, call 1-800-273-8255
- GLBT Youth Hotline, call 1-866-488-7386
- Beth Israel Deaconess Emergency Room, 190 Pilgrim Rd Boston, MA

Students requiring academic accommodations must provide an official accommodation memo from **The Center for Wellness** and contact me privately to discuss logistics.

COLLEGE OF THE FENWAY STUDENTS:

If you are enrolled in this course through COF Cross Registration, notify your course instructor. Please provide her/him with your email address to be sure that you receive course information in a timely way. You should also discuss how to access online applications that might be used in the course.

SYLLABUS OUTLINE:

WEEKLY SCHEDULE: A tentative schedule is provided below. The schedule is subject to change as the semester progresses.

Week	Topic	Assignments
1	Introduction to KDD (Knowledge Discovery in Databases) and Data Mining.	
2	Data Preprocessing	Assignment 1
3	Data Cube, Data Warehouse and Data Lake	
4	Data Mining algorithms: Frequent Itemset Mining	Assignment 2
5	Data Mining algorithms: Association Rule Mining	
6	Data Mining algorithms: Sequential Pattern Mining	
7	Classification in Data Mining	Assignment 3
8	Clustering in Data Mining	
9	Advanced topics: Text Mining	Assignment 4
10	Advanced topics: Multimedia Data Mining	
11	Advanced topics: Web Mining	
12	Ethics of Data Mining	Final Project
13	Final Project: How to write a scientific paper on Data Mining	Final Project
14	Final Project: How to write a scientific paper on Data Mining	Final Project
15	Final Project Presentations	Final Project

Othman, Salem is inviting you to a scheduled Zoom meeting.

Topic: Salem Othman's Zoom Meeting

Time: This is a recurring meeting Meet anytime

Join Zoom Meeting

https://wentworth.zoom.us/j/345383795?pwd=cEUzOXk4V3ZnM2FNVTZ6UVJBTllsZz09

Meeting ID: 345 383 795

Passcode: 002109

One tap mobile

+13052241968,,345383795# US

+13092053325,,345383795# US

Dial by your location

- +1 305 224 1968 US
- +1 309 205 3325 US
- +1 312 626 6799 US (Chicago)
- +1 646 558 8656 US (New York)
- +1 646 931 3860 US
- +1 301 715 8592 US (Washington DC)
- +1 346 248 7799 US (Houston)
- +1 360 209 5623 US
- +1 386 347 5053 US
- +1 507 473 4847 US
- +1 564 217 2000 US
- +1 669 444 9171 US
- +1 669 900 9128 US (San Jose)
- +1 689 278 1000 US
- +1 719 359 4580 US
- +1 253 205 0468 US
- +1 253 215 8782 US (Tacoma)

Meeting ID: 345 383 795

Find your local number: https://wentworth.zoom.us/u/kexycoCvrE