CptS 315: Introduction to Data Mining

Syllabus, Spring 2023

Course Information

Credit Hours: 3

Semester: Spring 2023

Meeting times and location: Tue and Thu, 9:10–10:25am, Bryan Hall 305.

Learning Management System: Piazza and Canvas will be used for the management of this course. Piazza will be used for posting of lecture material, assignments, announcements, and messages. Canvas will be use for student submissions, instructor/TA feedback, and posting grades.

Instructor and TA Information

Instructor: Jana Doppa

Office: EME 133

Email: jana.doppa@wsu.edu

Office Hours: Mon 4-5pm (EME 133)

Teaching Assistant: Chibuike Ugwu

Email: chibuike.ugwu@wsu.edu

Office Hours: Wed 2-3pm (Dana 122)

Teaching Assistant: Emma Mickas

Email: emma.mickas@wsu.edu

Office Hours: Fri 12-1pm (Dana 15)

Teaching Assistant: Anthony Ghimpu

Email: anthony.ghimpu@wsu.edu

Office Hours: Thu 1-2pm (Dana 136)

Course Description:

The course will cover various algorithms to automatically analyze data using computers for discovering knowledge and insights. For each data mining task, we will study three aspects:

1. Computational problems motivated from real-world applications
2. Computational (scalable) algorithms to solve data analysis problems
3. Real-world applications for this data mining task

Tentatively, we will cover the following topics:

- Mining frequent item sets and association rules
- Recommendation algorithms
- Supervised learning algorithms
    - Classification and regression tasks
- Data clustering algorithms
- Outlier and anomaly detection algorithms
- Computational advertising algorithms
- Algorithms for Mining Data Streams
- Responsible data mining

Prerequisites:

Assume strong programming experience (won't teach programming)

- You can choose any high-level programming language of your choice (e.g., Python, Java, C++)
- Strongly encourage you to consider Python (data analytics community is built around Python tools)

Assume knowledge of the following

- Basic data structures (CptS 223)
- Basic algorithms
- Time/space complexity analysis

Course Materials

- We will NOT follow a fixed textbook for this course
- Instructor will provide slides and lecture notes
- Slides and reading materials will be posted on Piazza site

Optional Textbooks

- [MMD] Mining of Massive Datasets: Jure Leskovec, Anand Rajaraman, and Jeff Ullman. Available online at
    - http://infolab.stanford.edu/~ullman/mmds/book0n.pdf
- [CIML] A Course in Machine Learning: Hal Daume
    - http://ciml.info/
- I will assign reading material from MMD and CIML

Student Learning Outcomes and Assessment Methods

Learning Outcomes

- Understand the challenges, algorithms, and applications of data mining tasks
- Scalability challenges and algorithmic solutions for scalable implementation
- Ability to apply basic data mining algorithms for solving real-world problems

Assessment Methods

- Homework assignments
- Programming assignments
- Exams
- Course project

Coursework and Grading Policies

- 3 Homeworks (30%)
- 2 Exams (40%)
    - Mid-term #1 and Mid-term #2
    - No exam during Finals week
- 1 Course Project (25%)

    - Individual or group of two students
- Class Participation (5%).
    - Easy part. Don't miss it!
    - Piazza and in-class
    - Attendance (Instructor will take attendance on 4 random days. To get attendance related grade, should be present in at least 2 occasions)

Late Policy

- All assignments, project proposal/report are due at midnight of the deadline day.
- Late Policy
    - 0-24 hours late -- 80% of the final score
    - 24-48 hours late -- 50% of the final score
    - Beyond 48 hours -- 0%

Grading Policy

- [100-93] A
- (93-90] A-
- (90-86] B+
- (86-83] B
- (83-80] B-
- (80-76] C+
- (76-73] C
- (73-70] C-
- (70-66] D+
- (66-60] D
- (60-0] F

Note: I may decide to move the thresholds down based on the distribution of final percentages.

Schedule for Due Dates:

HW1: Feb 9th midnight

Midterm Exam #1: Feb 23 (in class)

HW2: Mar 7th midnight

Course Project Proposal: Mar 23rd midnight

HW3: Mar 30th midnight

Midterm Exam #2: April 27 (in class)

Course Project Report: May 4th midnight

Complete Schedule:

Class	Day	Date	Topic	Notes
1	Tue	01/10/2023	Introduction and Logistics	
2	Thu	01/12/2023	Frequent Itemsets and Association Rules Mining	
3	Tue	01/17/2023	Frequent Itemsets and Association Rules Mining	
4	Thur	01/19/2023	Frequent Itemsets and Association Rules Mining HW1 Posted	
5	Tue	01/24/2023	Python libraries for data manipulation demo by TAs	
6	Thur	01/26/2023	Class canceled (Instructor was Sick)	
7	Tue	01/31/2023	Recommender Systems	
8	Thur	02/02/2023	Recommender Systems	

9	Tue	02/07/2023		
10	Thur	02/09/2023	HW1 Due	
11	Tue	02/14/2023		
12	Thur	02/16/2023	HW2 Posted	
13	Tue	02/21/2023		
14	Thur	02/23/2023	Midterm Exam #1	
15	Tue	02/28/2023		
16	Thur	03/02/2023		
17	Tue	03/07/2023	HW2 Due	
18	Thur	03/09/2023	HW3 Posted	
	Tue	03/14/2023	SPRING BREAK	
	Thur	03/16/2023	SPRING BREAK	
19	Tue	03/21/2023		
20	Thur	03/23/2023	Course Project Proposal Due	
21	Tue	03/28/2023		
22	Thur	03/30/2023	HW3 Due	
23	Tue	04/04/2023		
24	Thur	04/06/2023		
25	Tue	04/11/2023		
26	Thur	04/13/2023		
27	Tue	04/18/2023		
28	Thur	04/20/2023		
29	Tue	04/25/2023		

30	Thur	04/27/2023	Midterm Exam #2	
31	Tue	05/02/2023		
32	Thur	05/04/2023	Course Project Report Due	

Other Policies

Correspondence

All class related correspondence with the instructor will be made via Piazza

Academic Integrity

Academic integrity is the cornerstone of higher education. As such, all members of the university community share responsibility for maintaining and promoting the principles of integrity in all activities, including academic integrity and honest scholarship. Academic integrity will be strongly enforced in this course. Students who violate WSU's Academic Integrity Policy (identified in Washington Administrative Code (WAC) 504-26-010(4) will fail the assignment implicated, will not have the option to withdraw from the course pending an appeal, and will be reported to the Center for Community Standards.

Cheating includes, but is not limited to, plagiarism and unauthorized collaboration as defined in the Standards of Conduct for Students, WAC 504-26-010(3). Read and understand all of the definitions of cheating given here http://app.leg.wa.gov/WAC/default.aspx?cite=504-26-010. If you have any questions about what is and is not allowed, ask your course instructor.

If you wish to appeal an instructor's decision relating to academic integrity, please use the form available at http://communitystandards.wsu.edu. Make sure you submit your appeal within 21 calendar days of the instructor's decision.

Students with Disabilities

Reasonable accommodations are available for students with documented disabilities or chronic medical conditions. If you have a disability and need accommodations to fully participate in this class, please visit your campus Access Center website (websites listed below) to follow published procedures to request

accommodations. Students may also call or email the Access Center to schedule an appointment with an Access Advisor. All disability related accommodations are to be approved through the Access Center. It is a university expectation that students with approved accommodations visit with instructors (in person or via Zoom) within two weeks of requesting their accommodations to discuss logistics.

For more information contact a Disability Specialist on your home campus:

• Pullman or WSU Online: 509-335-3417 http://accesscenter.wsu.edu , Access.Center@wsu.edu

• Spokane: https://spokane.wsu.edu/studentaffairs/access-resources/

• Tri-Cities: http://www.tricity.wsu.edu/disability/

• Vancouver: 360-546-9138;

https://studentaffairs.vancouver.wsu.edu/student-wellness-center/access-center

Accommodation for Religious Observances or Activities

Washington State University reasonably accommodates absences allowing for students to take holidays for reasons of faith or conscience or organized activities conducted under the auspices of a religious denomination, church, or religious organization. Reasonable accommodation requires the student to coordinate with the instructor on scheduling examinations or other activities necessary for course completion. Students requesting accommodation must provide written notification within the first two weeks of the beginning of the course and include specific dates for absences. Approved accommodations for absences will not adversely impact student grades. Absence from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the period of absence. Students who feel they have been treated unfairly in terms of this accommodation may refer to Academic Regulation 104 – Academic Complaint Procedures. See also Academic Regulation 82, available at https://registrar.wsu.edu/academic-regulations/

Safety and Emergency Notification

Classroom and campus safety are of paramount importance at Washington State University and are the shared responsibility of the entire campus population. WSU

urges students to follow the “Alert, Assess, Act” protocol for all types of emergencies and “Run, Hide, Fight” response for an active shooter incident. Remain ALERT (through direct observation or emergency notification), ASSESS your specific situation, and act in most appropriate way to assure your own safety (and the safety of others if you are able).

Please sign up for emergency alerts on your account at MyWSU. For more information on this subject, campus safety and related topics, please view the FBI’s Run, Hide, Fight video (https://www.fbi.gov/about-us/cirg/active-shooter-and-mass-casualty-incidents/run-hide-fight-video) and visit the WSU safety portal (https://faculty.wsu.edu/classroom-safety). Full details can be found at https://provost.wsu.edu/classroom-safety/.

Academic Dates and Deadlines

Students are encouraged to refer to the academic calendar often to be aware of critical deadlines throughout the semester. The academic calendar can be found at http://registrar.wsu.edu/academic-calendar