CISC 483/683
Introduction to Data Mining

Credits: 3
Semester: Spring, Year: 2019

Meeting Days, Times, Location and Room: **TR 5:00 pm – 6:15 pm, 217 Gore Hall**

Course Home Page (Piazza)

Instructor: [Dr. Roghayeh (Leila) Barmaki](#), rlb@udel.edu, 406 Smith Hall

Teaching Assistants: **Debarati Chowdhury, and Haoran Wei**

Instructor Office Hours: Tuesdays 9:30 -10:30 am, 406 Smith Hall
TA Office Hours: **TBA**

Course Catalog Description: As an introductory course on data mining, this course introduces the concepts, algorithms, techniques, and systems of data mining, including (1) what is data mining? (2) get to know your data, (3) data preprocessing, integration and transformation, (4&5) mining frequent patterns and association: basic concepts and advanced methods, (6) classification: basic concepts, (7) cluster analysis: basic concepts, and current trends and special topics in data mining. The course will serve both senior-level computer science undergraduate students and graduate students interested in the field.

Prerequisites:

Background CISC-220 (data structures) and at least one upper-level course in computer science, or permission of instructor.
Programming We will give one or two programming assignments. You will need to be familiar with at least one programming language, such as C++, Java, or Python. We will not cover programming-specific issues in this course.

Textbook
Jiawei Han, Micheline Kamber and Jian Pei, Data Mining: Concepts and Techniques, **3rd ed**., Morgan Kaufmann, 2011.

See the book's [home page](#) for course slides and other reference materials.

Additional References
The following texts are recommended but not required, for reference. **There are numerous other books or online resources on data mining available.**

- E. Alpaydin. Introduction to Machine Learning, 2nd ed., MIT Press, 2011.
- T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer-Verlag, 2009.
- T. M. Mitchell, Machine Learning, McGraw Hill, 1997.
- P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison Wesley, 2005. Link.
- H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 2nd ed. 2005.
- Review of Probability Theory Tutorial - Stanford University

Lecture slides contains most technical briefing and reference materials. Please study the materials in class preparation and class review.

Course Format, Activities, and Evaluation
This course will draw materials mainly from the textbook (including the 3rd edition Textbook Draft), the course slides are important references. Students will study the materials and complete all the course requirements.

Reading: Before and After Classes
We encourage students to read ahead, before lectures for the materials to be discussed. Please check the **course schedule table** to see what will be covered in each lecture before the class begins.

Individual Written and Programming Assignments
There will be five to six assignments, spaced out over the course of the semester. Among these assignments, two will be programming assignments. **We recommend one of the programming languages of Java, C/C++, and Python for submitting programming assignments,** but if you have any concerns about these languages, please contact the TAs for help.

About programming assignments, it is OK to discuss the problems with the TAs and your classmates, however, it is NOT OK to work together or share code. Plagiarism is an academic violation to copy, to include text from other sources, including online sources, without proper citation.

Team Assignment: Final Report: Graduate Students (required), Undergrads (optional)
This will be a team assignment, required for graduates, and optional for undergraduates., The maximum team size for graduate student teams is 3. Undergraduate students can

also engage in the report writing (preferably with other undergraduate students in the teams of size 4 as maximum). See grading section for more details.

This report will be a comprehensive survey on a focused topic of data mining and machine learning, for example, a survey on clustering methods for heterogeneous information networks.

To form teams and brainstorm ideas, we will have a jumpstart session in the third week of class. More details will be provided later, but your reports need to follow the ACM double-column Format, and at least have 6 pages of length excluding references. All submitted reports will be checked via Turnitin.com for plagiarism. **We will have a Writing Fellow to help you to prepare your report.**

Finally, all of the teams will briefly present a summary of their work in a fast-forward session at the end of the semester.

Course Schedule (tentative)

Readings are from Han et al. Textbook.

#	Date	Readings/Topic	Remarks (tentative)
1	2/12 (Tue)	Class Outline / Chapter 1: Introduction	**HW 0 Out**
2	2/14 (Thu)	Chapter 2: Know Your Data	**HW 0 Due (11:59 pm)**
3	2/19 (Tue)	Chapter 2: Know Your Data (Additional Notes)	
4	2/21 (Thu)	Chapter 3: Data Preprocessing	
5	2/26 (Tue)	Chapter 3: Data Preprocessing	
6	2/28 (Thu)	Final Report Idea Session:	

		Forming Teams and Defining your Topics	
7	3/5 (Tue)	[Chapter 6: Mining Frequent Patterns and Associations: Basic Concepts](#)	
8	3/7 (Thu)	Chapter 6: Mining Frequent Patterns and Associations: Basic Concepts	
9	3/12 (Tue)	Chapter 6: Mining Frequent Patterns and Associations: Basic Concepts	
10	3/14 (Thu)	Chapter 6: Mining Frequent Patterns and Associations: Basic Concepts	
11	3/19 (Tue)	Guest Lecture: [Dr. Austin Brockmeier](#)	
12	3/21 (Thu)	Exam Review Session	
13	3/26 (Tue)	**Midterm Exam**	**Midterm Exam @ Class**
14	3/28 (Thu)	Midterm Exam Review, Probability Review	
–	4/2 (Tue)	**Spring Vacation**	
–	4/4 (Thu)	**Spring Vacation**	
15	4/9 (Tue)	[Chapter 8: Classification: Basic Concepts](#)	

16	4/11 (Thu)	Chapter 8: Classification: Basic Concepts	
17	4/16 (Tue)	Chapter 8: Classification: Basic Concepts	
18	4/18 (Thu)	Chapter 9: Classification: Advanced Concepts	
19	4/23 (Tue)	Chapter 10: Cluster Analysis: Basic Concepts	
20	4/25 (Thu)	Chapter 10: Cluster Analysis: Basic Concepts	
21	4/30 (Tue)	Chapter 10: Cluster Analysis: Basic Concepts	
22	5/2 (Thu)	Special Topics in Data Mining	
23	5/7(Tue)	Special Topics in Data Mining	
24	5/9 (Thu)	Chapter 13: Trends and Research Frontiers in Data Mining	
25	5/14 (Tue)	Final Report Presentations	
26	5/16 (Thu)	Final Exam Review	

Learning Resources

We will use both **Piazza and Canvas** for this course.

- For **course materials** (lecture notes, assignments, and other supplementary materials), **announcements, assignments, deliverables, and your grades**, we will use **Canvas**.

- **For asking questions** (**both public and private questions**) related to everything, including assignments and final report we will use **Piazza**. **Please do not post your code in Piazza.**

    o If you have concerns about your grade, send a private post via Piazza. If you have a general question, you can ask it from the entire class (see this Video).
    o Information on how to use Piazza is available through the Piazza Support Center.

- For
    o Information on how to use Canvas is available through the Canvas Student Guide.
    o Canvas is accessible via MyUD or directly at http://www.udel.edu/canvas.

Assessment of Learning*

We plan to determine final grades of the course in the following way:

Graduate Students:

Individual Assignments **15%** (4 homework assignments expected)

Programming assignments: **20%** (two programming assignments expected)
Midterm exam: **25%**
Final exam: **30%**
Team Final Report, and Fast-Forward Presentation **10% (Max Team size: 3)**

Undergraduate Students:

Individual Assignments **20%** (4 homework assignments expected)

Programming assignments: **20%** (two programming assignments expected)
Midterm exam: **25%**
Final exam: **30%**
Extra Credit: Team Final Report, and Fast-Forward Presentation **5% (Max Team size for undergrad-only group is 5)**

* Some of the assignments, and exams will be slightly different for graduate and undergraduate students. All of these criteria will be taken into account by the instructor in assigning each student's individual grade. we reserve the right to adjust the syllabus during the semester, and we will give you notice in class if we do so.

Class Participation
Class participation is strongly encouraged and leads to a much more enjoyable and productive class. So please actively contribute to the class discussions (both in the class and discussion boards) and feel free to ask questions. Particularly good class contributions will positively affect borderline decisions on final grades in the course. Disruptive or distracting behavior hurts the whole class; such behavior will result in a reduction of up to two letter grades (for example, from "A-" to "C-") in the student's final grade in the course.

Personal Devices & Laptops in Class
While use of laptops might be distracting for the entire class most of the time, we might ask you to bring and use your laptop for doing a few activities together in the class. If so, we will let you know to bring your laptop prior to that session.

Grading Scale
Students will be assigned the following letter grade based on the calculation coming from the course assessment section.

Grade	Interval
A	94.50 and over
A-	89.50 to 94.49
B+	86.50 to 89.49
B	82.50 to 86.49
B-	79.50 to 82.49
C+	76.50 to 79.49
C	72.50 to 76.49
C-	69.50 to 72.49
D+	66.50 to 69.49
D	62.50 to 66.49
D-	59.50 to 62.49
F	Below 59.49

Academic Integrity

Please familiarize yourself with UD policies regarding academic dishonesty. To falsify the results of one's research, to steal the words or ideas of another, to cheat on an assignment, to re-submit the same assignment for different classes, or to allow or assist another to commit these acts corrupts the educational process. Students are expected to do their own work and neither give nor receive unauthorized assistance. Complete details of the university's academic integrity policies and procedures can be found at http://www1.udel.edu/studentconduct/policyref.html Office of Student Conduct, 218 Hullihen Hall, (302) 831-2117. E-mail: student-conduct@udel.edu

Harassment and Discrimination

The University of Delaware works to promote an academic and work environment that is free from all forms of discrimination, including harassment. As a member of the community, your rights, resource and responsibilities are reflected in the non-discrimination and sexual misconduct policies. Please familiarize yourself with these policies at www.udel.edu/oei . You can report any concerns to the University's Office of Equity & Inclusion, at 305 Hullihen Hall, (302) 831-8063 or you can report anonymously through UD Police (302) 831-2222 or the EthicsPoint Compliance Hotline at www1.udel.edu/compliance. You can also report any violation of UD policy on harassment, discrimination, or abuse of any person at this site: sites.udel.edu/sexualmisconduct/how-to-report/

Faculty Statement on Disclosures of Instances of Sexual Misconduct

If, at any time during this course, I happen to be made aware that a student may have been the victim of sexual misconduct (including sexual harassment, sexual violence, domestic/dating violence, or stalking), I am obligated to inform the university's Title IX Coordinator. The university needs to know information about such incidents in order to offer resources to victims and to ensure a safe campus environment for everyone. The Title IX Coordinator will decide if the incident should be examined further. If such a situation is disclosed to me in class, in a paper assignment, or in office hours, I promise to protect your privacy--I will not disclose the incident to anyone but the Title IX Coordinator. For more information on Sexual Misconduct policies, where to get help, and how to reporting information, please refer to www.udel.edu/sexualmisconduct. At UD, we provide 24-hour crisis assistance and victim advocacy and counseling. Contact 302-831-1001, UD Helpline 24/7/365, to get in touch with a sexual offense support advocate.

For information on various places you can turn for help, more information on Sexual Misconduct policies, where to get help, and reporting information please refer to www.udel.edu/sexualmisconduct

Inclusion of Diverse Learning Needs

Any student who thinks he/she may need an accommodation based on a disability should contact the Office of Disability Support Services (DSS) office as soon as possible. The DSS office is located at 240 Academy Street, Alison Hall Suite 130, Phone: 302-831-

4643, fax: 302-831-3261, DSS website (www.udel.edu/DSS/). You may contact DSS at dssoffice@udel.edu

Non-Discrimination

The University of Delaware does not discriminate against any person on the basis of race, color, national origin, sex, gender identity or expression, sexual orientation, genetic information, marital status, disability, religion, age, veteran status or any other characteristic protected by applicable law in its employment, educational programs and activities, admissions policies, and scholarship and loan programs as required by Title IX of the Educational Amendments of 1972, the Americans with Disabilities Act of 1990, Section 504 of the Rehabilitation Act of 1973, Title VII of the Civil Rights Act of 1964, and other applicable statutes and University policies. The University of Delaware also prohibits unlawful harassment including sexual harassment and sexual violence.

For inquiries or complaints related to non-discrimination policies, please contact:
Director, Institutional Equity & Title IX Coordinator- Susan L. Groff, Ed.D. groff@udel.edu, 305 Hullihen Hall Newark, DE 19716 (302) 831-8063

For complaints related to Section 504 of the Rehabilitation Act of 1973 and/or the Americans with Disabilities Act, please contact: Director, Office of Disability Support Services, Anne L. Jannarone, M.Ed., Ed.S. - ajannaro@udel.edu
Alison Hall, Suite 130, Newark, DE 19716 (302) 831-4643 OR contact the U.S.
Department of Education - Office for Civil Rights
(wdcrobcolp01.ed.gov/CFAPPS/OCR/contactus.cfm)