Data Mining
CS699 A1, Fall 2019

- **Course Format**: On Campus
- **Time and Location**: Tuesday 6:00 – 8:45 PM, PHO 211

- **Instructor**: Jae Young Lee
- **Office**: Room 303, 1010 Commonwealth Ave.
- **Phone**: 617-358-5165, **E-mail**: jaeylee@bu.edu
- **Office Hours**: 3:30 – 5:00 PM, Monday and Tuesday, and by appointment

- **Course Objectives**

  The goal of this course is to study basic concepts and techniques of data mining. The topics include data preparation, classification, performance evaluation, association rule mining, clustering, and time series analysis. We will discuss basic data mining algorithms in the class and students will practice data mining techniques using data mining software. Students will use Weka and JMP Pro or Oracle.

- **Prerequisites**:
  - CS 546 and either CS 579 or CS 669. Or instructor's consent.

- **Text**: Jiawei Han, M. Kamber, and J. Pei, "Data Mining Concepts and Techniques," Third Ed., 2012, Morgan Kaufmann

- **References**:
  - Ian H. Witten, E. Frank, and M.A. Hall, "Data Mining Practical Machine Learning Tools and Techniques," Fourth Ed., 2016, Morgan Kaufmann
  - G. Shmueli, P.C. Bruce, M.L. Stephens, N.R. Patel, 'Data Mining for Business Analytics; Concepts, Techniques, and Applications with JMP PRO," 2017, Wiley

- **Courseware**: Blackboard

- **Grading**:
  - Midterm: 25%, Final: 35%
  - Homework Assignment: 30%
  - Class Project: 10%

- **Letter Grade**:

  $90 \leq G < 94$: A- $94 \leq G$: A,
  $80 \leq G < 83$: B- $83 \leq G < 87$: B $87 \leq G < 90$: B+
  $70 \leq G < 73$: C- $73 \leq G < 77$: C $77 \leq G < 80$: C+
  $60 \leq G < 70$: D
  $G < 60$: F

  Note: Course grades are not automatically rounded up. For example, if your course grade is 93.9, you will receive A-, not A.

- **Assignment**
  - There will be 9 homework assignments (the number of homework assignments is subject to change).
  - Solutions will be discussed in the class.
  - Should be submitted on the Blackboard unless other submission method is specified in the assignment.

- **Class Project**:
  - This is a data mining project. Details will be discussed in the class later.
  - Should be submitted on the Blackboard.
  - All students must present their projects.

- **Midterm**: In-class, closed book, closed notes
- **Final**: In-class; closed book; any printed material, including hand-written notes, is allowed **except the textbook**; comprehensive; no electronic devices allowed.

- **Academic Integrity Policy**
  - Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions.
  - Please take the time to review the Student Academic Conduct Code: http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.html.
  - This should not be understood as a discouragement for discussing the material or your particular approach to a problem with other students in the class. On the contrary – you should share your thoughts, questions and solutions. Naturally, if you choose to work in a group, you will be expected to come up with more than one and highly original solutions rather than the same mistakes.
- **Attendance and Absence**:
  - Attendance is not required but strongly encouraged. If a student misses a class it is their responsibility to catch up with the material discussed during the missed class.

- **Late Policy**
  - All assignments are due at the beginning of the class on the due date.
  - There is a 2-day grace period. If you submit your assignments after the grace period, there will be a late penalty of 10% per day. An assignment is not accepted after 6 days from the assigned day.

- **Make-up Exam**
  - A make-up examination can be arranged for the midterm only when a student has an emergency (e.g., a medical emergency or an urgent family matter). Students may need to provide the instructor with an appropriate document (such as a letter from a physician).
  - There will be **no make-up exam for the final exam**. If a student cannot take the final exam on the designated day, they will receive an incomplete grade.

Tentative Schedule

- The schedule is subject to change according to the actual progress of the class. Some topics may be skipped and some topics may be added.
- We will also discuss some topics that are not in the textbook. These topics will be included in lecture slides which will be posted on Blackboard.

Week	Date	Topics	Textbook	Assignments (assigned date)	Project
1	9/3	Introduction to data mining, Introduction to data warehousing, Software installation	Chapter 1, Section 4.1	HW1	Project assigned
2	9/10	Data exploration	Sections 2.1, 2.2, 2.4.1, 2.4.2, 2.4.3, 2.4.4, 2.4.7	HW2	
3	9/17	Data preparation	Sections 3.1, 3.2, 3.3, 3.4.1, 3.4.6, 3.4.7, 3.4.8, 3.5	HW3	Proposal due
4	9/24	Classification 1	Sections 8.1, 8.3	HW4	
5	10/1	**Midterm**			
6	10/8	Classification 2 Performance evaluation	Sections 8.2.1, 8.2.2, 8.2.3, 8.4.1, 8.4.2, Section 8.5	HW5	Progress report due
7	10/15	**No class**			
8	10/22	Performance evaluation, Other classifiers	Sections 8.5, Section 8.6, Section 9.2, Section 9.5.1	HW6	
9	10/29	Association rule mining	Sections 6.1, 6.2.1, 6.2.2	HW7	
10	11/5	Pattern evaluation measures, Other association analysis	Sections 6.3, 7.2.1, 9.4.1	HW8	Project due
11	11/12	Clustering	Sections 10.1, 10.2.1, 10.3.1, 10.3.2, 10.6	HW9	
12	11/19	Selected topics			
13	11/26	Presentation			
14	12/3	Presentation			
15	12/10	Presentation			
16	12/17 (?)	**Final Exam**			

- **Communication**
  - All official announcements will be made in the class.
  - All assignments will be posted on the class web page.
  - **Important:** The primary method of communication is through in-class announcements. The class web page is only supplementary.
  - **Email communication**: When it is necessary to communicate to you, I will send an email to your BU email account. So, you need to check your BU email regularly (e.g., once a day).