Data Mining
CS699 A1, Spring 2025

- **Course Format**: On Campus
- **Time and Location**: Monday 6 – 8:45 PM, CAS 218

- **Instructor**: Jae Young Lee
- **Office**: Room 303, 1010 Commonwealth Ave.
- **Phone**: 617-358-5165, **E-mail**: jaeylee@bu.edu
- **Office Hours**:
  - 3:30 – 4:30 PM Tuesday and Thursday, and by appointment
  - Students can meet me in person (in my office) or via zoom
  - No office hours during the exam period

- **Course Description**

  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, and clustering. We will discuss basic data mining algorithms in the class and students will practice data mining techniques using R.

- **Prerequisites**:
  - CS544 or knowledge of R, or instructor's consent.

- **Text (required)**: Galit Shmueli et al., "Machine Learning for Business Analytics: Concepts, Techniques, and Applications in R," Second Ed. 2023, Wiley

- **Reference (recommended)**: Max Kuhn, Kjell Johnson, "Applied Predictive Modeling," 2nd Printing, Springer, 2018

- **Courseware**: Blackboard

- **Grading**:
  - Midterm: 25%, Final: 35%
  - Homework Assignment: 20%
  - Class Project: 20%

- **Letter Grade**:

  $90 \le G < 94$: A- $94 \le G$: A,
  $80 \le G < 83$: B- $83 \le G < 87$: B $87 \le G < 90$: B+
  $70 \le G < 73$: C- $73 \le G < 77$: C $77 \le G < 80$: C+
  $60 \le G < 70$: 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 10 homework assignments (the number of homework assignments is subject to change).
  - 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.

- **Exams**
  - Both the midterm and the final exams are in-class, paper-based exams.
  - Both exams are closed-book exams.
  - The final exam is a comprehensive exam.
  - Details will be discussed in the class.

- **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 study the material discussed during the missed class.

- **Late Policy**
  - All assignments are due at the beginning of the class on the due date.
  - If you submit an assignment late, a penalty of 10% per day will be imposed.
  - If a student obtains a permission from the instructor or the TA **in advance**, a late penalty may be waived.

- **Make-up Exam**
  - A make-up examination for the midterm can be arranged only when a student has an emergency (e.g., a medical emergency or an urgent family matter). Students must contact the instructor **before the exam** and may need to provide 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, she/he 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	Homework Assignment	Project Assignment
1	1/27	Introduction, Data Exploration	HW1	Assigned
2	2/3	Data exploration, Data Preprocessing	HW2	
3	2/10	Data preprocessing	HW3	
4	2/18 **Tuesday**	Classification 1	HW4	Intermediate Report due
5	2/24	Performance evaluation	HW5	
6	3/3	**Midterm**		
7	3/10	**No class**		
8	3/17	Classification 2, Intervention	Hw6	
9	3/24	Association rule mining, Collaborative filtering	HW7	
10	3/31	Pattern evaluation measures, Other association analysis	HW8	
11	4/7	Clustering	HW9	Final Report due
12	4/14	Time Series Analysis	HW10	Slides due
13	4/23 **Wednesday**	Presentation		
14	4/28	Presentation		
15		**Final Exam**		

- **Software tool:**
  - We will primarily use R.

- **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, at least once a day.
  - When you send an email to me, include "CS699 A1" in the subject of your email.