Data Mining
CS699 SC1, Summer 2025

- **Time and location**: Thursday 6:00 – 9:30 PM, CAS B06A

- **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 Wednesday and Thursday, via Zoom or in person in my office

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

- **Prerequisites**:
  - Knowledge of R, or instructor's consent

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

- **References**:
  - Galit Shmueli et al., "Machine Learning for Business Analytics: Concepts, Techniques, and Applications in R," Second Ed. 2023, Wiley.
  - 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**:
  $90 \leq G < 94$: 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

- **Assignment**
  - There will be 9 homework assignments (the number of homework assignments is subject to change).
  - Should be submitted on the Blackboard.

- **Class Project**:
  - This is a data mining project. Details will be discussed in the class.
  - This is a team project; 2 students per team
  - Should be submitted on the Blackboard.

- **Midterm**: In-class, paper-based exam; details will be discussed later
- **Final exam**: Comprehensive; in-class, paper-based exam; details will be discussed later

- **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, they are responsible for the material discussed during the class.

- **Late Policy**
  - All assignments are due at the beginning of the class on the due date.
  - There is a late penalty of 10% per day. An assignment is not accepted 6 days after the due date.

- **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 may need to provide the instructor with appropriate documents (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.

Week	Date	Topics	Assignments (assigned date)
1	5/22	Introduction, Data exploration	HW1
2	5/29	Data preparation	HW2
3	6/5	Classification 1	HW3
4	6/12	Performance evaluation	HW4 (not graded)
5	6/19	**No class**	
6	6/26	**Midterm**	
7	7/3	Classification 2	HW5
8	7/10	Association rule mining	HW6
9	7/17	Pattern evaluation, Other association analysis	HW7
10	7/24	Clustering	HW8
11	7/31	Introduction to time series analysis	HW9 (not graded)
12	8/7	**Final Exam**	

Software tool: We will use R.

Communication:

- When it is necessary to communicate with 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 SC1" in the subject of your email.