Data Mining
CS699 A2, Fall 2024

- **Course Format**: On Campus
- **Time and Location**: Thursday 6 – 8:45 PM, HAR 322

- **Instructor**: Jae Young Lee
- **Office**: Room 303, 1010 Commonwealth Ave.
- **Phone**: 617-358-5165, **E-mail**: jaeylee@bu.edu
- **Office Hours**:
  - 3 – 4 PM Tuesday and Thursday, and by appointment
  - Students can meet me in person (in my office) or via zoom
  - No office hours during the final exam week (12/16 – 12/20)

- **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**:
  - 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**:

  $90 \leq G < 94$: 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).
  - 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 **in advance**, a late penalty will be waived.

- **Make-up Exam**
  - A make-up examination 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 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	Book Chapter	Assignments
1	9/5	Introduction, Data exploration	C1, C2, C3	H1
2	9/12	Data exploration, Data preprocessing	C3, C4	H2
3	9/19	Performance evaluation	C5	H3
4	9/26	Multiple regression, KNN	C6, C7	
5	10/3	**Midterm**		
6	10/10	Naïve Bayes, Decision tree	C8, C9	H4
7	10/17	Logistic regression, SVM, ANN	C10, C11	H5
8	10/24	Discriminant analysis, Ensemble methods	C12 C13	H6
9	10/31	Intervention, Association rule mining, Collaborative filtering	C14 C15	H7
10	11/7	Clustering	C16	H8
11	11/14	Time series analysis	C17, C18 C19	H9
12	11/21	Presentation		
13	**11/28**	**No class**		
14	12/5	Presentation		
15	**TBD**	**Final Exam**		

- **Software tool:**
  - We will use R.
  - If, however, you do not have any experience/knowledge in R, you have the following options:
    - You can try to learn R and use R for assignments.
    - If you don't want to learn and use R, then let me know and I will prepare separate assignments for you, which do not require R.
    - Details will be discussed in the class.
  - If you have taken CS544 and/or CS555 or if you are taking CS544 and/or CS555 this semester, you **must 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 A2" in the subject of your email.