Data Mining
CS699D1/CS699EL, Fall 2015

- **Course Format**: On Campus / Blended
- **Time and Location**: Thursday 6:00 – 9:00 PM, FLR 267
- **699EL Section meets on 9/3, 10/1, 11/5, 12/10, and the final exam.**

- **Instructor**: Jae Young Lee
- **Office**: Room 250, 808 Commonwealth Ave.
- **Phone**: 617-358-5165, **E-mail**: jaeylee@bu.edu

- **Office Hours**: 4:00 – 5:30 PM, Monday and Thursday

- **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. Students will use Weka and SQL Server or Oracle.

- **Prerequisites**:
  - CIS students: CS546 and CS669
  - CS students: CS579
  - 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," Third Ed., 2011, Morgan Kaufmann

- **Courseware**: Blackboard Learn, URL: https://learn.bu.edu

- **Grading**:
  - Midterm: 30%, Final: 40%
  - Homework Assignment: 20%
  - Class Project: 10%

- **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		

- **Assignment**
  - There will be 10 homework assignments.
  - Solutions will be discussed in the class when graded papers are returned.
  - Should be submitted on the Blackboard unless other submission method is specified in the assignment.

- **Class Project**:
  - This is a data mining project. It will be assigned on 10/15 and will be due on 11/19.
  - Should be submitted on the Blackboard.

- **Midterm**: In-class, closed-book
- **Final**: In-class, open-book (any printed material is allowed except the textbook), comprehensive

- **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 his/her responsibility to catch up with the material discussed during the missed class.
  - 699EL students must attend all four class meetings plus the in-class final exam.

- **Late Policy**
  - All assignments are due at the beginning of the class on the due date.
  - A late assignment is subject to a penalty of 10% per day. An exception may be made if a student is in an unusual/urgent situation and obtains permission from the instructor before the due date.

- **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.

- **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 neither in the textbook nor in the online lecture. These topics will be included in lecture slides which will be posted on Blackboard.

Week	Date	Topics	Textbook	Online Lecture	Homework (assigned)
1	9/3	Introduction to data mining, Introduction to data warehousing, Weka, SQL Server, Oracle	Chapter 1, Section 4.1	Module 1	HW1
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	Module 2	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	Module 2	HW3
4	9/24	Classification, 1R, Naïve Bayes, Decision tree	Sections 8.1, 8.2.1, 8.2.2, 8.2.3, 8.3	Module 3	HW4
5	10/1	Rule based classifier, Classification with Weka	Sections 8.4	Module 3	HW5
6	10/8	**Midterm**			Project
7	10/15	Performance evaluation	Sections 8.5.1, 8.5.2, 8.5.3, 8.5.4, 8.5.5, 8.5.6	Module 4	HW6
8	10/22	Other classifiers	Sections 8.6.1, 8.6.2, 8.6.3, Slides	Module 4	HW7
9	10/29	Association rule mining	Sections 6.1, 6.2.1, 6.2.2	Module 5	
10	11/5	Correlation analysis, Other association analysis	Sections 6.3.1, 6.3.2, 6.3.3, 7.2.1, 9.4.1	Module 5	HW8
11	11/12	Clustering	Sections 10.1, 10.2.1, 10.3.1, 10.3.2, 10.6	Module 6	HW9
12	11/19	Clustering	Slides	Module 6	HW10 Project due
13	11/26	**No class**			
14	12/3	Case studies	Slides		
15	12/10	Selected topics	Slides		
16		**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).