Syllabus

This is a single, concatenated file, suitable for printing or saving as a PDF for offline viewing. Please note that some animations or images may not work.

Course Description

This module is also available as a concatenated page, suitable for printing or saving as a PDF for offline viewing.

MET CS699

Data Mining

Data mining is a one of the most important components in the knowledge discovery process. The course provides an introduction to concepts and techniques behind data mining. The course surveys various data mining applications, methodologies, techniques, and models. Topics include classification, association rules, and clustering. Algorithms will be tested on data sets using the Weka Data mining software and Microsoft SQL Server or Oracle.

The course grading will consist of analyzing a series of assignments, weekly quizzes, a project, and an open book, proctored final exam.

Course Overview

- Overview of Data Mining and Data Warehousing
- Getting Started with SQL Server 2012 or Oracle and WEKA Tools
- Data Preparation
- Classification
- Association Rule Mining
- Clustering
- Case Studies

Technical Note

The table of contents expands and contracts (+/- sign) and may conceal some pages. To avoid missing content pages, you are advised to use the next/previous page icons in the top right corner of

the learning modules.

Course Learning Objectives

By successfully completing this course you will be able to:

- Understand what Data Mining is and where it is applicable
- Distinguish Data Mining from Data Warehousing, OLAP, and Query Processing
- Study Data Mining techniques, including Classification, Association Rule Mining, Clustering, etc.
- Apply practical examples using WEKA and SQL Server or Oracle

Course Outline

- **Calendar Tool** - You can see due dates in the calendar tool. You may add your own events there as well. However, please be aware that you may not find all of the important dates for the course listed there. You will stay current by checking announcements, discussions, and emails in the course.
- **Readings** - Each module has both textbook readings and online lectures . Your professor may suggest additional readings during the running of the course.
- **Discussion** - There are group-level threaded discussions for each module. These discussions are moderated by your facilitator. Postings for each discussion should be completed by the assigned due dates. There are also course-level general discussions boards, which are not graded, for you to use to discuss any issues with your classmates. Please see the Discussion Module on the home page for more details.
- **Assignment** - There are assignments that are due throughout the course. They are accessed from the Assignments menu.
- **Assessments** - Weekly quizzes and final exams are also listed in the course calendar and accessed from the Assessments menu.

Module 1 - Introduction and Software Installation

- Overview of Data Mining and Data Warehousing
- Getting Started with SQL Server or Oracle and WEKA

Module 2 - Data Exploration and Preprocessing

- Data Exploration
- Data Preparation

Module 3 – Classification

- Naïve Bayes
- Decision Tree

Module 4 – Performance Evaluation and Other Classifiers

- Performance Evaluation
- Other Classifiers

Module 5 – Association Analysis

- Association Analysis
- Association Rule Evaluation

Module 6 – Clustering and Case Studies

- Clustering
- Case Studies

Module 7 – Prepare for and take the final exam

You will prepare for and take the proctored final exam.

The course will remain open two weeks after the final exam, so that you can continue discussions and ask any questions about database technology, your grades, or the course. This is also a time when we enter into a dialog where we endeavor to learn from you how we can modify the course so that it better meets students' needs.

Instructor

Jae Young Lee, Ph.D.

Computer Science Department
Metropolitan College
Boston University
808 Commonwealth Ave
Boston, MA 02215

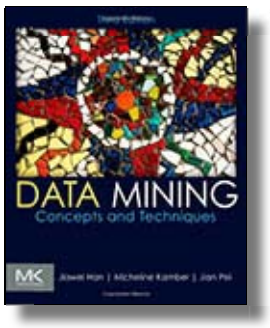
Phone: 617-358-5165
Email: jaeylee@bu.edu
Live Office hours: TBA

Professor Jae Young Lee received his Ph.D. in Computer Science from the Computer Science and Engineering Department of the University of Texas at Arlington. He joined the Metropolitan College of Boston University in the fall of 2007, and has been teaching various computer science courses, including Database Management, Database Administration, Data Mining and Business Intelligence, Analysis of Algorithms, and Artificial Intelligence. Before joining BU, he taught at the Colorado School of Mines and the University of North Florida.

His research areas include conceptual modeling, query language, association rule mining, and outlier detection. Recently, he has been studying how to efficiently detect outlier (or unusual) sequence from a large sequence dataset.

Course Materials

Required Book



Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* - 3rd Edition.
Morgan Kaufmann.

ISBN-13: 978-0123814791
ISBN-10: 0123814790

This book can be purchased from Barnes and Noble at Boston University.

Recommended Book

Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* - 4th Edition.
Morgan Kaufmann.

ISBN-13: 978-0128042915

ISBN-10: 0128042915

For reference only and **not required** to be purchased for this course.

This book can be purchased from Barnes and Noble at Boston University.

Supplemental Material

Download the following zip file containing the supplementary materials to your local hard disk: Supplements

Virtual Lab

Metropolitan College is a member of The VMware Academic Program. VMAP enables current MET students and faculty to gain easy access to cutting-edge virtualization technology and resources.

All current MET students are granted access to download. For information on how to login and get support, please visit: http://www.bu.edu/metit/hw-and-sw/vmware-academic-program/

MathJax

Variables, formulae, and equations in this course are rendered using MathJax.

ing Fractions

ıg fractions: $\frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm cb}{bd}$, often this is
:asy to remember. a, b, c, d do not have t

To enable its features in your browser, right-click (or ctrl-click on a single-mouse-button Mac) on a variable or equation to see your MathJax settings.

MathJax can be used with the MathPlayer plugin for Internet Explorer, which converts math to speech and highlights the math as it is spoken.

Boston University Library Information

Boston University has created a set of videos to help orient you to the online resources at your disposal. An

introduction to the series is below:

met_ode_library_14_sp1_00_intro is displayed here

Download

All of the videos in the series are available on the Online Library Resources page, which is also accessible from the Ca.m.pus Bookmarks section of your Online Ca.m.pus Dashboard. Please feel free to make use of them.

As Boston University students, you have full access to the BU Library. From any computer, you can gain access to anything at the library that is electronically formatted. To connect to the library, use the link http://www.bu.edu/library. You may use the library's content whether you are connected through your online course or not, by confirming your status as a BU community member using your Kerberos password.

Once in the library system, you can use the links under "Resources" and "Collections" to find databases, eJournals, and eBooks, as well as search the library by subject. Some other useful links follow:

Go to http://www.bu.edu/library/research/collections to access eBooks and eJournals directly.

If you have questions about library resources, go to http://www.bu.edu/library/help/ask-a-librarian to email the library or use the live-chat feature.

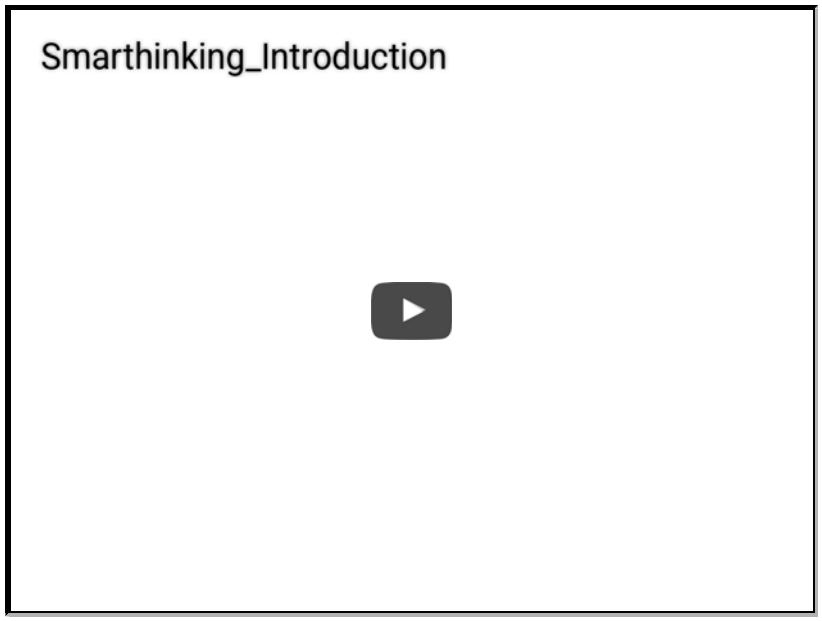
To locate course eReserves, go to http://www.bu.edu/library/services/reserves.

Please note that you are not to post attachments of the required or other readings in the water cooler or other areas of the course, as it is an infringement on copyright laws and department policy. All students have access to the library system and will need to develop research skills that include how to find articles through library systems and databases.

Free Tutoring Service

Free online tutoring with SMARTHINKING is available to BU online students for the duration of their courses. The tutors do not rewrite assignments, but instead teach students how to improve their skills in the following areas: writing, math, sciences, business, ESL, and Word/Excel/PowerPoint.

You can log in directly to SMARTHINKING from Online Campus by using the link in the left-hand navigation menu of your course.



Please Note

The SMARTHINKING service can be used for Boston University online class work only. Use of this service for personal purposes or for anything other than Boston University online class work will result in deactivation of your SMARTHINKING account.

Study Guide

Module 1 Study Guide and Deliverables

Readings:	*Online Lectures* *Data Mining: Concepts and Techniques:* Chapter 1 — Introduction Chapter 4 – Data Warehouse and OLAP, Basic Concepts: Section 4.1
Discussions:	Discussion 1 postings end Tuesday, May 16 at 6:00 a.m. ET
Assignments:	Assignment 1 due Tuesday, May 16 at 6:00 a.m. ET
Assessments:	Quiz 1 due date Tuesday, May 16 at 6:00 a.m. ET
Live Classrooms:	Live Classroom Wednesday, May 10 at 7:30 p.m. ET Live Q&A Session Saturday, May 13 at 11:00 a.m. ET

Module 2 Study Guide and Deliverables

Readings:	Online Lectures *Data Mining: Concepts and Techniques:* Chapter 2 –Getting to Know Your Data: Sections 2.1, 2.2, 2.3, 2.4.1, 2.4.2, 2.4.3, 2.4.4, 2.4.5, 2.4.7 Chapter 3 – Data Preprocessing: Sections 3.1, 3.2, 3.3, 3.4.1, 3.4.6, 3.4.7, 3.4.8, 3.4.9, 3.5
Discussions:	Discussion 2 postings end Tuesday, May 23 at 6:00 a.m. ET
Assignments:	Assignment 2 due Tuesday, May 23 at 6:00 a.m. ET
Assessments:	Quiz 2 due Tuesday, May 23 at 6:00 a.m. ET
Live Classrooms:	Live Classroom Thursday, May 18 at 7:30 p.m. ET Live Q&A Session Saturday, May 20 12 at 11:00 a.m. ET

Module 3 Study Guide and Deliverables

Readings:	Online Lectures *Data Mining: Concepts and Techniques* Chapter 8 – Classification:Basic Concepts: Sections 8.1, 8.2.1, 8.2.2, 8.2.3, 8.3, 8.4.1, 8.4.2
Discussions:	Discussion 3 postings end Tuesday, May 30 at 6:00 a.m. ET
Assignments:	Assignment 3 due Tuesday, May 30 at 6:00 a.m. ET
Assessments:	Quiz 3 due Tuesday, May 30 at 6:00 a.m. ET
Live Classrooms:	Live Classroom Thursday, May 25 at 7:30 p.m. ET Live Q&A Session Saturday, May 27 19 at 11:00 a.m. ET

Module 4 Study Guide and Deliverables

Readings:	Online Lectures *Data Mining: Concepts and Techniques* Sections 8.5.1, 8.5.2, 8.5.3, 8.5.4, 8.5.5, 8.6.1, 8.6.2, 9.5.1
Discussions:	Discussion 4 postings end Tuesday, June 6 at 6:00 a.m. ET
Assignments:	Assignment 4 due Tuesday, June 6 at 6:00 a.m. ET
Assessments:	Quiz 4 due Tuesday, June 6 at 6:00 a.m. ET

Live Classrooms: Live Classroom Thursday, June 1 at 7:30 p.m. ET
Live Q&A Session Saturday, June 3 at 11:00 a.m. ET

Module 5 Study Guide and Deliverables

Readings: Online Lectures
Data Mining: Concepts and Techniques
Chapter 6 – Association Analysis: Sections 6.1, 6.2.1, 6.2.2, 6.3.1, 6.3.2, 6.3.3

Discussions: Discussion 5 postings end Tuesday, June 13 at 6:00 a.m. ET

Assignments: Assignment 5 due Tuesday, June 13 at 6:00 a.m. ET

Assessments: Quiz 5 due Tuesday, June 13 at 6:00 a.m. ET

Live Classrooms: Live Classroom Thursday, June 8 at 7:30 p.m. ET
Live Q&A Session Saturday, June 10 at 11:00 a.m. ET

Module 6 Study Guide and Deliverables

Readings: Online Lectures
Data Mining: Concepts and Techniques
Chapter 10 Cluster Analysis - Sections 10.1.1, 10.1.2, 10.1.3, 10.2.1, 10.3.1, 10.3.2, 10.6

Discussions: Discussion 6 postings end Tuesday, June 20 at 6:00 a.m. ET

Assignments: Assignment 6 and Project due Tuesday, June 20 at 6:00 a.m. ET

Live Classrooms: Live Classroom Thursday, June 15 at 7:30 p.m. ET
Live Q&A Session Saturday, June 17 at 11:00 a.m. ET

Final Exam Details

The Final Exam is a proctored exam available from **June 21 at 6:00 a.m. ET to June 24 at 11:59 p.m. ET**. The Computer Science department requires that all final exams be proctored.

The exam is a three-hour open-book exam consisting of 20 multiple-choice questions, and you are allowed to bring a calculator. It will only be accessible during the final exam period. You can access it from the Assessments Menu of the course. Your proctor will enter the password to start the exam.

You will receive a technical support hotline number before the start of the exam. Please bring this number with you to the exam.

Course Grading Information

Grading Structure and Distribution

The grade for the course is determined by the following:

Overall Grading Percentages	
Assignments	30
Discussions	10
Project	10
Quizzes	20
Proctored Final Examination	30

94 ≤ G	A
90 ≤ G < 94	A−
87 ≤ G < 90	B+
83 ≤ G < 87	B
80 ≤ G < 83	B−

77 ≤ G < 80	C+
73 ≤ G < 77	C
70 ≤ G < 73	C−
60 ≤ G < 70	D
G < 60	F

Grades will be curved to maintain academic standards at Boston University.

Live Classroom Discussion and Archive

The instructor of the course will be conducting five synchronous Live Classroom discussion in Weeks 1–6, and they will be archived for further viewing.

Your participation, while not mandatory, will be valuable to you and the class. To participate in the discussion or to access the archived session, you will need to go to the Live Classroom link on your homepage.

You must have speakers and a microphone for your computer. A headset is recommended. If you choose to connect via phone, enter the live classroom and click on the telephone icon. Use the phone number and passcode provided.

Discussions

Graded Discussions—Students will be participating in discussions that will be graded on a 100-point scale: go to the Discussion Rubric

Quizzes

There will be five quizzes, one per every week except in Week 6 and Week 7. The primary goal of quizzes is to let students keep current with the course material.

Proctored Final Exam

There will be a proctored Final Exam for this course. You will be responsible for scheduling your own appointment with an approved proctoring option. Detailed instructions about setting up an appointment will be forthcoming from the proctored exam coordinator.

Expectations

It is important for each student to participate on a regular basis and complete all aspects of this course. This course is designed to include a major portion of learning by interacting (asynchronously) with the other students in the class, and the grade is therefore dependent on this activity.

Delays

If, for any reason, you are unable to meet any deadline, contact your Course Facilitator. Assignments are expected to be submitted by their respective due dates. Extensions may be granted, though **only under mitigating circumstances**. If your facilitator grants an extension, you will not be penalized. If you submit an assignment late without a permission of your facilitator, there will be a late penalty of 10% per day.

Discussion Grading Rubric

Graded discussion periods are held Day 1 (Tuesday) until 6:00 AM ET on Day 1 of the following week. You're certainly welcome to continue a discussion past the grading period, but that additional posted material will not affect your discussion grade. The discussion grading rubric below is the guide we use to evaluate your discussion contributions.

Criteria	65–69	70–79	80–89	90–94	95–100
Participation	Very limited participation	Participation generally lacks frequency or relevance	Reasonably useful relevant participation during the discussion period	Frequently relevant and consistent participation throughout the discussion period	Continually relevant and consistent participation throughout the discussion period
Community	Mostly indifferent to discussion	Little effort to keep discussions	Reasonable effort to respond	Often responds thoughtfully in	Continually responds thoughtfully in a

		going or provide help	thoughtfully, provide help, and/or keep discussions going	a way that frequently keeps discussions going and provides help	way that consistently keeps discussions going and provides help
Content	No useful, on-topic, or interesting information, ideas, or analysis	Hardly any useful, on-topic, or interesting information, ideas, or analysis	Reasonably useful, on-topic, and interesting information, ideas, and/or analysis	Frequently useful, on-topic, and interesting information, ideas, and analysis	Exceptionally useful, on-topic, and interesting information, ideas, and analysis
Reflection and Synthesis	No significant effort to clarify, summarize, or synthesize topics raised in discussions			Contributes to group's effort to clarify, summarize, or synthesize topics raised in discussions	Leads group's effort to clarify, summarize, or synthesize topics raised in discussions

Weekly Quiz Instructions

Accessing the Quiz

You will have access to the quiz at the beginning of the week. However you should not access the quiz until you have completed all learning activities for the week and are prepared to meet the objectives for that week.

Quiz Details

- There are 20 questions per quiz. You can access the quiz details from the assessments menu.
- The questions are either multiple choice or True/False.
- All questions are randomized.
- The points for each question are shown.
- The quiz questions will display one at a time on your screen.

- You may skip over questions and revisit them in any order.
- You will have 90 minutes to complete the quiz. You should have enough time so that you aren't rushed. Also note:
- You can take each quiz only once.
- You may not pause the quiz and return to it later.
- You will be able to continue to save answers to questions after the time has expired, but any late answers will be time stamped and marked as late. This will allow us to grade your quiz fairly in the event that technical difficulties occur while you take your quiz.

Saving Answers

- To answer a multiple choice question, select the appropriate choice from the list below the question.
- When you have completed your response, click "Save Answer" at the top of the question.
- As you proceed through the exam, you can go back and edit previous responses that you saved.
- A timer is displayed above the questions tracking the remaining time available.
- You will see question number buttons above questions. You will need to click on "Question Completion Status" to see the question numbers. You can use these buttons to navigate from question to question at any time.
- When you have completed all answers, go to the last question of the exam and click the "Save and Submit" button.

If a technical issue of any kind arises during the quiz requiring you to go beyond the time limit, complete the quiz answering the remaining questions and then contact your facilitator or instructor immediately.

Comments on the Quiz

There will be a short-answer area at the end of the quiz; it appears as a quiz question, but there are no points for this item. Use this as a place to provide feedback about the quiz as a whole or to comment upon a particular quiz item. Be sure to reference the question number. Your facilitator will examine your comments in order to decide whether a grade adjustment or other action should be taken.

If a technical issue of any kind arises during the exam, complete the exam, answering the remaining questions, and then contact your facilitator or instructor immediately.

Other Questions

If you have any questions about the quiz please feel free to contact your facilitator.

Technical Support

Assistance with course-related technical problems is provided by the IS&T Help Center. To ensure the fastest possible response, please fill out the online form using the link below.

IT Help Center Support
888-243-4596 or local 617-353-4357 or Web
Check your open tickets using BU's ticketing system.

Boston University Metropolitan College