

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 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 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 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 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

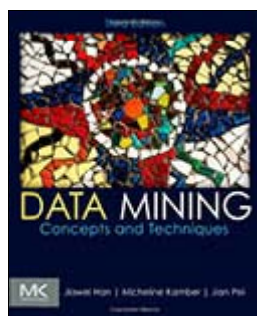


Professor Jae Young Lee received his Ph.D. in Computer Science from the Computer Science and Engineering Department of the University of Texas at Arlington. He joined the Metropolitan College of Boston University in the fall of 2007, and has been teaching various computer science courses, including Database 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](#).