

MIS 6500: Data Mining

Instructor & Course Information:

Instructor: Yong Seog Kim, PhD
Phone: (435) 797-2271
E-Mail: yong.kim@usu.edu
Office: BUS 710

Course Location: HH 124
Course Times: TR 1:30-2:45
Course Website: <http://canvas.usu.edu>
Office Hours: TR 9:15-10:15

Introduction

It is impossible to ignore the importance of data mining and business intelligence techniques in the contemporary severe competitive environment. Data mining and business intelligence techniques have been actively applied to many business domains to identify problems and provide solutions in a timely manner. A few typical problems that can be solved using data mining techniques include churn prediction (banking and marketing application), cross-selling (retailing application), fraud detection (banking and accounting application), risk management (banking and finance application), customer segmentation (banking, marketing, and retailing application), targeted ads (retailing application), and sales forecast (all business domains).

This course provides an introduction to data mining and business intelligence technologies. This course will start with a brief but sufficient theoretical introduction of data mining and business intelligence. Then, students will learn how to evaluate and visualize the performance of data mining algorithms. In particular, two of three representative services of Microsoft SQL 2016 Server, SQL Server Integration Services (SSIS) and SQL Server Analysis Services (SSAS), will be explored in detail to study several well-known data mining algorithms for clustering, association rule mining, and classification purposes. These tools will be also used to visualize the outcomes of such algorithms and draw business insights. Additional tools such as Weka and R Studio will be introduced to study other popular data mining algorithms such as ensemble models, text mining models, and data reductions techniques if time permits.

The detailed objectives of this class are mapped to the IDEA center course evaluation objectives used by USU as shown below. **Please note that you will be asked to rate this course at the end of the semester based on the IDEA Center objectives listed.**

Course Objectives

Upon completion of this course, students should master following skills:

IDEA Center Learning Objectives	MIS 6500 Learning Objectives
Objective 1 (Essential): Gaining factual knowledge (terminology, classifications, methods, trends) Objective 2 (Important): Learning fundamental principles, generalizations, or theories	1. Understand how to evaluate and visualize the performance of data mining algorithms. 2. Understand underlying assumptions and algorithm details of classification techniques. 3. Understand underlying assumptions, parameters settings, and algorithm details of clustering techniques.

	<p>4. Understand underlying assumptions, parameters settings, and algorithm details of association rule mining techniques.</p> <p>5. Understand underlying assumptions, parameters settings, and algorithm details of text mining techniques.</p>
<p>Objective 3 (Essential): Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course</p> <p>Objective 4 (Important): Learning to <i>apply</i> course material (to improve thinking, problem solving, and decisions)</p>	<p>1. Develop and implement various classification algorithms such as ZeroR, naïve Bayesian, logistic regression, neural network, decision tree, and support vector machines.</p> <p>2. Develop and implement clustering algorithms such as K-means and EM and study the relationship between the number of clusters and clustering outputs.</p> <p>3. Develop and implement an association rule mining algorithm to identify a set of items that are frequently purchased together.</p> <p>4. Develop and implement a text mining to represent actively exchanged text patterns in a word cloud from social network sites.</p> <p>5. Visualize the performance of various data mining models and draw new marketing and other business insights.</p> <p>6. Visualize key attributes in data sets using Tableau.</p>

Required and Recommended Textbooks

The recommended textbooks for the course are the following:

1. Title: Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking
Author: Foster Provost and Tom Fawcett
Publisher: O'Reilly Media (2013). ISBN 978-1449361327. 1st edition
Required: No

2. Title: Data Mining: Practical Machine Learning Tools and Techniques
Author: Ian Witten, Eibe Frank, and Mark Hall
Publisher: Morgan Kaufmann (2011). ISBN 978-0-12-374856-0. 3rd edition
Required: No

Prerequisites

Students should not only have strong understanding in database and data warehouse (MIS 6230 and MIS 6110-02) but also be able to install and maintain two main data mining tools: MS SQL Server 2016 and Weka.

Course Materials/Access

I will use Canvas (<http://canvas.usu.edu>) to post course material, including lecture notes, assignments, grades, etc. It is very important that you access Canvas early and regularly in the semester, since you will be responsible any course material, change, or notification that is posted there.

Required Software

Students are required to install SQL Server Enterprise 2016 server module on their own computers. Additional software such as Weka, Tableau, and R Studio may be introduced and used if necessary. The course fee for MIS 6500 (\$30) is used to provide lab support, hardware maintenance and software licensing support.

Course Approach

MIS 6500 takes a holistic approach which includes textbook readings, faculty presentations, computer-based instruction, and practical exercises. The textbook readings and faculty presentations will convey the main body of knowledge for this class and include a variety of terms, concepts, and issues related to Data Mining and Business Intelligence. The computer-based instruction will demonstrate ways of using various complex tools (e.g., SSDT and SSMS) of SQL Server 2016. Finally, the exercises and assignments that require data mining tools will provide students the opportunity to be familiar with SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS), Tableau, Weka, and R Studio.

Grading Policy

Final grades will be determined as follows:

Test 1	25%
Test 2	25%
Assignments	25%
Data mining group project	15%
Class Activity	6%
Individual Presentation	2%
Professional conduct	1%
Attendance & class participation	1%

A	A-	B+	B	B-	C+	C	C-	D+	D	F
≥93%	≥90%	≥87%	≥83%	≥80%	≥77%	≥73%	≥70%	≥67%	≥60%	<60%

Exams (50%)

There will be one midterm (25%) and one final exam (25%) which will consist of multiple-choice, fill-in-blank, and/or short essay questions that cover the material presented in your textbook and lecture notes. **It is crucial that you read your textbook and lecture notes to perform well on the exams**, as some material may not be covered during class.

Assignments (25%)

There will be several assignments that require comprehensive understand of course materials including text book exercises, lecture notes, and additional time outside of class for your own studies. These assignments must be:

1. Submitted ON TIME. No late work is accepted. Due dates for each assignment will appear on the Canvas course calendar. Occasionally, there may be technical problems with Canvas that prevent you from submitting your assignment. In such cases, you may email your assignment to me in advance. However, I must receive the email **on or before the assignment due date** or no credit will be given.
2. Completed individually by you and you alone. You are welcome to ask other students for assistance, but your work must be your own.
3. Completed professionally (see Professionalism section above).

Data mining group project (15%)

Students will be required to work on a data mining project that analyzes a sizable data set to demonstrate their understanding of data mining algorithms and evaluation metrics. Forming and working with a great team is one of the most important factors to determine whether your team project will succeed or fail, which will naturally lead to your final grade. Groups will be formed in the class and specific information will be given. Each team may include two or three members. Due dates will be noted on the Canvas.

Class activity (6%)

Students will be required to work on several class activities that fresh their understanding on data mining theories and algorithms as a group. Groups will be formed in the class and specific information will be given. Each team may include two or three members. Due dates will be noted on the Canvas.

Individual presentation (2%)

Each student will be required to present a short review on a data mining algorithm or business impact of data mining on a specific business area such as finance, marketing, human resources management and so on. Topics for each student will be determined in the class and due dates will be noted on the Canvas.

Professional conduct (1%)

Professionalism implies positive participation in class discussions and an appropriate attitude for learning. I request that you always demonstrate professionalism in this course, in your work as well as in your conduct. I expect that you will demonstrate professionalism in your work by:

1. Submitting your work on time. All assignments are typically due by the midnight but you are allowed to upload your work by 5AM of the following day without any penalty. Remember you can submit an incomplete version and replace it with new one. Always check if you upload the final version. I will deduct 0.5% from your professional conduct credit (2%) each time when you ask for an extension of the due date of assignments for following or similar reasons
 - you (including your babies, dogs, and cats) accidently delete or corrupt your file
 - you forget to submit your work on time due to an annual family union event or fatigue
 - you submit an incomplete or wrong version of your work (check if it is the final)
 - you have slow or lost Internet connection around the midnight on due date (plan ahead)
 - you could not complete assignment because your computer was suddenly broken, stolen, or malfunctioned for hardware or software reasons.

2. Displaying integrity by doing your own work; **never plagiarize the work of others**, including the textbook.
3. Submitting college graduate-level work. Your writing should consist of complete sentences, and should be free from spelling and grammatical errors. It should demonstrate to me that you understand the material and that you can thoughtfully justify your answers.

Attendance and class participation (1%)

I expect that you will demonstrate your class participation by:

1. Coming to class regularly and on time. Although I may not take roll every class and I understand if you need to miss class periodically, I expect that you will attend class consistently.
2. Participating in class discussions and exercises.
3. Do not engaging in activities that show disrespect to me or to your fellow students, including talking/texting on cell phones or browsing the Internet during class.
4. The Utah State University General Bulletin states that: "A student is expected to attend all meetings of a class for which he or she is registered. A student may be dropped from a course by the Dean if absences are repeated and the instructor recommends this action. A student can gain readmission only with permission of both the Dean and the instructor. A student dropped from a course receives an 'F' which counts as work attempted whenever grade point ratio calculations are made. Students with absences in excess of 10% of scheduled class meetings may be administratively withdrawn from the class with a grade of F which will count as work attempted whenever grade point ratio calculations are made."

Alternative Format Available for Course Materials

Students with disabilities are encouraged to discuss their needs with the instructor, preferably during the first week of the semester. Course materials in alternative format large print, audio, and Braille will be provided in cooperation with the Disability Resource Center (Ext. 7-2444). Reasonable accommodation will be provided for all persons with disabilities to ensure equal participation in class.

Comfortable Learning Environment

USU, the MIS Department, and your instructor are all committed to maintaining an inoffensive, non-threatening learning environment for every student. Class members (including the instructor) are thus to treat each other politely both in word and deed. Offensive humor and aggressive personal advances are specifically forbidden. If you feel uncomfortable with a personal interaction in class, see your instructor for help in solving the problem. The MIS Department Head, the Dean of the Huntsman School of Business, and USU's Affirmative Action Office are also willing to help as needed.

Incompletes

An "I" (Incomplete) grade is not to be used for avoiding a bad grade – this is University policy. Only extraordinary and unforeseen circumstances that prohibit a student from completing the course are adequate reason for giving an Incomplete grade.

Cheating

There is no room whatsoever for cheating in this class. Anyone caught cheating in **any way** will at best receive a failing grade for the course and at worst will be kicked out of the university. We have no

tolerance for cheating. What is cheating? Cheating may include (but is not limited to): plagiarizing the work of other students or material from the textbook, Web, or other sources; utilizing more resources than allowed while completing exams (including other students, books, the Web, or other materials); and, finally, turning in the work of another student (present or past) as your own work.