Course description: Data mining refers to exploration of data to discover knowledge. The knowledge discovered goes beyond the general pattern finding where queries are known. Data mining is an analytic process to discover the unknowns about the data. The topics covered in this course are useful to gain insight and expertise on mining large-scale datasets. Along with regular lectures and discussions in this course, there will be a semester-long project and hands-on activities, especially on algorithmics, tool development, and data analysis. The course will also cover recent state-of-the-art algorithms used to discover useful information from massive amount of data. This course is beneficial for students both from industry and research perspectives.

Instructor: Mahmud Shahriar Hossain  
Email: mhossain [at] utep [dot] edu  
Web: http://www.cs.utep.edu/mhossain/  
Phone: (915) 747 6340  
Office: CCSB 3.0504  
Office Hours: Tuesdays and Thursdays (1:00 pm - 2:00 pm) and by appointment

Class meetings: Tuesdays and Thursdays (10:30 am - 11:50 am)  
Meeting place: Health Science/School of NURS 213 or Classroom Building C204  
Prerequisite: Data structures, and at least one statistics course  
Reference books:  
• Data Mining: Concepts and Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei  
• Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar  
• Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, http://nlp.stanford.edu/IR-book/

EVALUATION:  
Grading components:  
Midterm 15%  
Final 15%  
Homework, quiz, in-class exercise 25%  
Project 40%  
Attendance 5%

The grading scale is A: 90-100, B: 80-89, C: 70-79, D: 60-69, F: below 60.
Grade appeals: All exam/homework/quiz grades must be appealed within 7 days of the grade being posted.

Project: There will be a semester long group project. Students are encouraged to form groups of two to three members. Each group can decide to work on a topic that closely relates to the interest of the members. The project must have major data mining components in it. The target of the project is to deliver a publishable report along with the description of the methodology and full experimental results. It is expected that the final report of the project will be submitted to a major data mining conference or a relevant workshop.

Exams: A midterm and final exam will be given. Make-up exams will not be permitted except under unusual circumstances with satisfactory written justification. Any student who misses an exam due to an unexcused absence will receive a grade of zero for that exam with no opportunity for make-up or substitution. University excused absences will be excused; the exam related arrangements should be made in advance in those cases.

Homework: Regular homeworks will be assigned which will require significant effort outside of class. The assignments are designed to challenge you by requiring that you apply learned concepts to new situations. You should start your homework immediately after you receive it.

Quizzes and exercises: There will be regular quizzes and exercises in the class. The quizzes are not scheduled rather may appear suddenly in any day. There will be individual exercises in the class as well.

Attendance: The instructor's policy is to penalize those students who are absent. Students are expected to actively participate in classes, and show the courtesy by not arriving late or leaving early. Although attendance has a weight of 5% of the total score, the instructor reserves the right to penalize the final grade for low attendance based on the fact that collecting information regarding technology and active participation in the classroom environment is the core of this course.

CLASS POLICIES:

Electronic devices: If you bring cell phones, or similar electronic equipment into the classroom, please turn them off or put them in a "quiet" mode. Please do not read text messages or send text messages during the class. It is ok to bring a laptop in the classroom but you cannot chat, check messages, or surf on the internet unless the instructor advises you to do so.

Wearing headphones of any type is strictly prohibited in the classroom. Headphones can be permitted only if the student has appropriate documentation approved by the Center for Accommodations and Support Services (CASS).

Standards of Conduct: You are expected to conduct yourself in a professional and courteous manner, as prescribed by the UTEP Standards of Conduct (http://sa.utep.edu/osccr/student-conduct/, http://studentaffairs.utep.edu).

Graded work, e.g., homework and tests, is to be completed independently and should be unmistakably your own work (or, in the case of group work, your team's work), although you may discuss your project with other students in a general way. You may not represent as your own work material that is transcribed or copied from another person, book, or any other source, e.g., a web page.

Academic dishonesty includes but is not limited to cheating, plagiarism and collusion.
• Cheating may involve copying from or providing information to another student, possessing unauthorized materials during a test, or falsifying data (for example program outputs) in laboratory reports.
• Plagiarism occurs when someone represents the work or ideas of another person as his/her own.
• Collusion involves collaborating with another person to commit an academically dishonest act.

Professors are required to -- and will -- report academic dishonesty and any other violation of the Standards of Conduct to the Dean of Students (http://sa.utep.edu/dean/).

DISABILITIES:
If you have a disability and need classroom accommodations, please contact The Center for Accommodations and Support Services (CASS) at 747-5148, or by email to cass@utep.edu, or visit their office located in UTEP Union East, Room 106. For additional information, please visit the CASS website at www.sa.utep.edu/cass.

TOPICS:
This course will cover the basic techniques in data mining including the preparation and manipulation of data for analysis and the creation of data from multiple dissimilar sources. Students will get hands-on experience on the data mining and knowledge discovery process.

1. Introduction
2. Essentials of Data Mining
3. Data Preprocessing
4. Data Warehousing and Online Analytical Processing
5. Data Cube Technology
6. Mining Frequent Patterns, Associations, and Correlations
7. Classification
8. Cluster Analysis
9. Outlier Detection
10. Advanced Topics

The instructor will include topics that he finds relevant as the semester progresses.

OTHER REQUIREMENTS:
Students are encouraged to bring a laptop in the classroom for in-class exercises and hands-on activities. We will be using Matlab’s Statistical Toolbox, Optimization Toolbox, Global Optimization Toolbox, and Parallel Computing Toolbox for some of the exercises. It is recommended that students have Matlab Student version installed in their machines.

The instructor reserves the right to make necessary changes to this syllabus and to the delivery of the course.