Introduction to Data Mining

Course description

This course will introduce fundamental and advanced strategies and methodologies for data mining along with the concepts underlying them, and will provide hands-on experience with a variety of different techniques ranging from statistics to machine learning to database. In addition, the MapReduce technique is introduced for big data mining. Students will learn to use the Weka workbench, a set of data mining tools.

Learning Objectives

Upon successful completion of this course, a student will
- comprehend the essential principles of contemporary data mining techniques;
- be able to apply data mining techniques to analyze large real world data sets;
- be familiar with data preprocessing and cleaning techniques associated with data mining;
- Understand the MapReduce technique in big data mining.

Major Course Requirements

The course grade will be determined as follows:

- Project and term research paper: 45%
- Research paper presentation: 15%
- Exams (1): 25%
- Class attendance: 5%
- Class participation: 10%

Required or Recommended Readings

Prerequisite: COSC 5335 and COSC 5321

Recommended or Supplemental Reading:
You might also find the following useful (it is the companion book to WEKA, which will be used for course projects):

Course Policies

Attendance/tardiness
The students are expected to come to class on time every day the class meets. Read the chapter to be discussed before coming to class. Ask questions of material you do not understand. If I cannot explain the answers to your satisfaction, make an appointment with me to discuss the question. Demonstrate integrity, maturity, and ethical behavior.

Late work and Make-up Exams
Assignments are accepted until MIDNIGHT on the due date. Every homework assignment will list a due date for full credit. Late assignments will lose 10% of the maximum score per day. Makeup exams will not be given under normal circumstances. If you notify me immediately that serious, unavoidable, documentable (e.g., with a letter from your doctor) circumstances have arisen, I will discuss options for replacing the missing grade. (For example, I may allow the grade earned on the comprehensive final to replace the grade for the missed exam.) Excused absences due to school sponsored activities, religious observations, family rituals, etc. should be discussed in advance.

**Academic Integrity/Plagiarism.**

University students are expected to conduct themselves in accordance with the highest standards of academic honesty. Academic misconduct for which a student is subject to penalty includes all forms of cheating, such as illicit possession of examinations or examination materials, falsification, forgery, complicity or plagiarism. (Plagiarism is the presentation of the work of another as one’s own work.) In this class, academic misconduct or complicity in an act of academic misconduct on an assignment or test will result in the fail of the class.

**Dropping a Class**

I hope that you never find it necessary to drop this or any other class. However, events can sometimes occur that make dropping a course necessary or wise. Please consult with me before you decide to drop to be sure it is the best thing to do. Should dropping the course be the best course of action, you must initiate the process to drop the course by going to the Student Services Center and filling out a course drop form. Just stopping attendance and participation WILL NOT automatically result in your being dropped from the class. 11/02/2012 is the last day to drop a class with an automatic grade of “W” this term.

**Classroom/professional behavior**

Texas A&M University-Corpus Christi, as an academic community, requires that each individual respect the needs of others to study and learn in a peaceful atmosphere. Under Article III of the Student Code of Conduct, classroom behavior that interferes with either (a) the instructor’s ability to conduct the class or (b) the ability of other students to profit from the instructional program may be considered a breach of the peace and is subject to disciplinary sanction outlined in article VII of the Student Code of Conduct. Students engaging in unacceptable behavior may be instructed to leave the classroom. This prohibition applies to all instructional forums, including classrooms, electronic classrooms, labs, discussion groups, field trips, etc.

**Grade Appeals**

A student who believes that he or she has not been held to appropriate academic standards as outlined in the class syllabus, equitable evaluation procedures, or appropriate grading, may appeal the final grade given in the course. The burden of proof is on the student to demonstrate the appropriateness of the appeal. A student with a complaint about a grade is encouraged to first discuss the matter with the instructor. For complete details on the process, including the responsibilities of the parties involved in the process and the number of days allowed for completing the steps in the process, consult Texas A&M University-Corpus Christi University Procedure 13.02.99.C2.01 Student Grade Appeal Procedures (http://www.tamucc.edu/provost/university_rules/index.html), and the College of Science and Engineering Grade Appeals webpage (http://sci.tamucc.edu/students/GradeAppeal.html). For assistance and/or guidance in the grade appeal process, students may contact the chair or director of the appropriate department or school or the College of Science and Engineering Dean’s Office.

**Disabilities Accommodations**

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If
you believe you have a disability requiring an accommodation, please call or visit Disability Services at (361) 825-5816 in Driftwood 101.

If you are a returning veteran and are experiencing cognitive and/or physical access issues in the classroom or on campus, please contact the Disability Services office for assistance at (361) 825-5816.

**Syllabus**

*(course outline)*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Reading Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>Chapter 1</td>
</tr>
<tr>
<td>Getting to Know Your Data</td>
<td>Chapter 2</td>
</tr>
<tr>
<td>Data Preprocessing</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>Mining Frequent Patterns, Associations</td>
<td>Chapter 6</td>
</tr>
<tr>
<td>and Correlations</td>
<td></td>
</tr>
<tr>
<td>Classification: Basic Concepts</td>
<td>Chapter 8</td>
</tr>
<tr>
<td>Cluster Analysis: Basic Concepts</td>
<td>Chapter 10</td>
</tr>
<tr>
<td>Exam I: Chapters 1, 2, 3, 6, 8, 10</td>
<td></td>
</tr>
<tr>
<td>Classification: Advanced Methods</td>
<td>Chapter 9</td>
</tr>
<tr>
<td>Advanced Cluster Analysis</td>
<td>Chapter 11</td>
</tr>
<tr>
<td>Data Mining for Big Data Analysis</td>
<td></td>
</tr>
<tr>
<td>Term Paper Presentation</td>
<td></td>
</tr>
</tbody>
</table>