COURSE SYLLABUS

Semester: Spring 2013

Course Prefix/Number: CAP5771

Course Title: Data Mining

Course Credit Hours: 3.0

Course Meeting Times/Places:
Online

Instructor and Contact Information:
Dr. Dallas Snider
Office: Building 4, Room 233
E-mail: dsnider@uwf.edu
Tel: 850-473-7348

Office Hours: Monday, Wednesday and Friday 9:00am – 11:00am; and by appointment.

Course Web Site: http://elearning.uwf.edu/ (login and select Data Mining, CAP5771)

Prerequisites or Co-requisites: Database Systems (COP4710 or COP5725).

Course Description:
This course will expose students to data mining concepts and techniques, and data mining software. Briefly, the topics that will be covered are: data preprocessing and cleaning, concept hierarchy generation, attribute relevance analysis, association rule mining and decision tree induction.

Student Learning Outcomes:
Upon successful completion of this course, the student will be able to:
- Identify data mining functionalities
- Identify data warehousing functionalities
- Apply data preprocessing techniques - data cleaning, data integration and transformation, data reduction, discretization, and concept hierarchy generation
- Describe data mining primitives, languages, and system architectures
- Use a Data Mining Query language
- Apply analytical characterization techniques
- Perform attribute relevance analysis
- Mine descriptive statistical measures in large databases
- Mine association rules in large databases
- Use classification by decision tree induction techniques on large databases
- Use data mining software to perform data mining functionalities – mining association rules, classification and prediction and decision tree analysis.

Topics Covered:
What is Data Mining, what is data warehousing, data mining functionalities – what kinds of patterns can be mined, what is data warehousing, data preprocessing – data cleaning, data integration and transformation, data reduction, discretization and concept hierarchy generation, data mining primitives, languages, and system architectures, attribute analysis,
mining class comparisons, association rule mining, classification by decision tree induction, issues in contemporary data mining applications.

**Required Textbook(s):**  
*Data Mining: Concepts and Techniques, Third Edition* by Jiawei Han, Micheline Kamber and Jian Pei, 2012. ISBN: 978-0-12-381479-1.

**References:**  

**Tentative grade distribution:**  
- Quizzes: 15%  
- Midterm: 15%  
- Final exam: 20%  
- Assignments: 20%  
- Project: 30%

**Technology Requirements:**  
Knowledge of a machine learning tool – WEKA (on the Windows environment) will be necessary for the project.

**Expectations for Academic Conduct/Plagiarism Policy:**  
Academic Conduct Policy: [(Web Format)](#) | [(PDF Format)](#) | [(RTF Format)](#)  
Plagiarism Policy: [(Word Format)](#) | [(PDF Format)](#) | [(RTF Format)](#)  
Student Handbook: [(PDF Format)](#)

**Assistance:**  
Students with special needs who require specific examination-related or other course-related accommodations should contact the Director of Disabled Student Services (DSS), [dss@uwf.edu](mailto:dss@uwf.edu), (850) 474-2387. DSS will provide the student with a letter for the instructor that will specify any recommended accommodations.

**Attendance:**  
Attendance is very strongly encouraged. Students are responsible for all announcements and all material covered in class. If you miss a class, you do not have to email me to let me know that you will be absent, just please ask a classmate for notes to determine what you missed.

**Exams:**  
Makeup exams will NOT be given except with a serious, documented medical or legal excuse. No makeups will be given unless students make advance arrangements. The final exam is comprehensive.

**Re-grading Assignments:**  
It is the student’s responsibility to check graded assignments/tests when they are returned to you. I will gladly re-grade an assignment/test when a question or mistake is brought to my attention. To ensure fairness, I reserve the right to re-grade the entire assignment/test. As a result, your grade may increase, decrease, or remain the same. Grades will not be changed after a week from the date graded assignments/tests are returned to the class.

**Grades:**  
Final grades will be calculated using a standard grade distribution. The last day of the term for withdrawal from an individual course with an automatic grade of “W” is **March 22**. Students requesting **late withdrawal** (W or WF) from class (deadline for this is **April 26**), must have the approval of the advisor, instructor, and department chairperson (in that
Requests for late withdraws may be approved only for the following reasons (which must be documented):

1. A death in the immediate family.
2. Serious illness of the student or an immediate family member.
3. A situation deemed similar to categories 1 and 2 by all in the approval process.
4. Withdrawal due to Military Service (Florida Statute 1004.07)
5. National Guard Troops Ordered into Active Service (Florida Statute 250.482)

Requests without documentation will not be accepted. Requests for late withdrawal simply for not succeeding in a course, do not meet the criteria for approval and will not be approved.

Applying for an incomplete or “I” grade will be considered only if: (1) there are extenuating circumstances to warrant it, AND (2) you have a passing grade and have completed at least 70% of the course work, AND (3) approval of the department chair.

**Participation and Feedback:** I encourage active participation in class. I believe that effective communication between the instructor and students will make the course more useful, interesting, and productive. **Please contact me if you have any questions, concerns, or suggestions! 😊**

**Social Media Policy:**
In order to provide fairness and appropriate professional separation between instructor and students, I will not accept social media invitations from students.

**Important Note:** *Any changes to the syllabus or schedule made during the semester take precedence over this version.* Check the eLearning site (or email) regularly for up-to-date information.

**Overall Grading Scale:**
- A: 93 - 100
- A-: 89 – 92.999
- B+: 87 – 88.999
- B: 82 – 86.999
- B-: 79 – 81.999
- C+: 77 – 78.999
- C: 72 – 76.999
- C-: 69-71.999
- D+: 67-68.999
- D: 59-66.999
- F: 0-58.999

**Late Policy:**
1. You are expected to complete work on schedule. Deadlines are part of the real world environment you are being prepared for.
2. Documentation of health or family problems may be required.
3. If you have to miss a class, be sure you arrange with another student to find out what you missed at the earliest possible date.
4. Late assignments will not be accepted.

*There’s another page…keep scrolling down…*
Tentative Course Schedule:

<table>
<thead>
<tr>
<th>WEEK #</th>
<th>WEEK OF</th>
<th>TOPIC</th>
<th>PROJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 07</td>
<td>Chapter 1 – Introduction, What is data mining?</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Jan 14</td>
<td>Chapter 2 – Getting to know your data</td>
<td>Datasets distributed</td>
</tr>
<tr>
<td>3</td>
<td>Jan 21</td>
<td>Chapter 3 – Data Preprocessing</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Jan 28</td>
<td>Chapter 4 – Data Warehousing and Online Analytical Processing</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Feb 04</td>
<td>Chapter 6 – Mining Frequent Patterns, Associations, and Correlations/ Presenting DM software</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Feb 11</td>
<td>Chapter 8 – Classification: Basic Concepts</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Feb 18</td>
<td>Chapter 9 – Classification: Advanced Methods</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Feb 25</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Mar 04</td>
<td>Chapter 10 – Cluster Analysis</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Mar 11</td>
<td>SPRING BREAK</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Mar 18</td>
<td>Chapter 12 – Outlier detection</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Mar 25</td>
<td>Chapter 13 – Data mining trends and research frontiers</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Apr 01</td>
<td>Data Mining Applications</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Apr 08</td>
<td>Project Q and A, time to work on project</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Apr 15</td>
<td>Project Q and A, time to work on project</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Apr 22</td>
<td>Review for final exam</td>
<td>PowerPoint presentations due</td>
</tr>
</tbody>
</table>

**Important Note:** Assignments will be assigned throughout the semester. Due dates for assignments will be given in class. Also, there is a major project in this class. The project will be a semester project that will have many parts. As the semester progresses, I will give you milestones when each part of the project will be due. Students are expected to do all parts of the project and all parts of the project will receive a grade (adding to up 30% in the end).

Also, tests and quizzes will be announced as the semester progresses.

Note for graduate students: Graduate students will be assigned additional work – please get in touch with the instructor (by the end of the 3rd week) if you are a graduate student.

*Enjoy the semester…*

*Dr Snider*