Introduction to Data Mining COSC-5590.002
School of Engineering and Computing Sciences
Summer I 2017

A. COURSE INFORMATION

Course number/section: COSC-5590.002
Class meeting time: TR 12:00-2:30PM
Class location: CS-111
Course Website: bb9.tamucc.edu

B. INSTRUCTOR INFORMATION

Instructor: Dr. Longzhuang Li
Office location: CI-323
Office hours: M 10:00-12:00PM
TR 2:30-4:00PM
Telephone: 825-2406
e-mail: Longzhuang.Li@tamucc.edu
Appointments: Call me or send me an email

C. COURSE DESCRIPTION

This course will introduce fundamental 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, with a focus on analysis of large data sets. Students will learn to use the Weka workbench, a set of data mining tools.

Catalog Description 3 semester hours
An introduction to fundamental strategies and methodologies for data mining. Topics include data preprocessing, mining frequent data patterns, classification, clustering, and outlier detection. Summer

D. PREREQUISITES AND COREQUISITES

Prerequisites
COSC 5335 and COSC 5321

Corequisites
None

E. REQUIRED TEXTBOOK(S), READINGS AND SUPPLIES

Optional Textbook(s) or Other References

Supplies
None

F. STUDENT LEARNING OUTCOMES AND ASSESSMENT
Assessment is a process used by instructors to help improve learning. Assessment is essential for effective learning because it provides feedback to both students and instructors. A critical step in this process is making clear the course’s student learning outcomes that describe what students are expected to learn to be successful in the course. The student learning outcomes for this course are listed below. By collecting data and sharing it with students on how well they are accomplishing these learning outcomes students can more efficiently and effectively focus their learning efforts. This information can also help instructors identify challenging areas for students and adjust their teaching approach to facilitate learning.

- Understand the basic statistical description of data, and be able to measure data similarity and dissimilarity.
- Be able to handle missing and noisy data.
- Understand the frequent itemset mining methods.
- Be able to analyze data using different classification methods, including decision tree, Bayes classification methods, rule-based classification.
- Understand and use a variety of cluster analysis methods.
- Understand outlier detection methods.

By the end of this course, students should be able to:
1. Comprehend the essential principles of contemporary data mining techniques;
2. Be familiar with data preprocessing and cleaning techniques associated with data mining;
3. Be able to apply data mining techniques to analyze large real world data sets.

G. INSTRUCTIONAL METHODS AND ACTIVITIES
The class includes lectures, presentations, term papers, and term projects. Students are expected to actively participate in the discussion.

H. MAJOR COURSE REQUIREMENTS AND GRADING
Grade Scale: A (90-100%)  B (80-89%)  C (70-79%)  D (60-69%)  F (<60%)

<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>% of FINAL GRADE</th>
</tr>
</thead>
</table>

2
I. COURSE CONTENT/SCHEDULE

<table>
<thead>
<tr>
<th>WEEK</th>
<th>TOPIC</th>
<th>CHAPTER(S)</th>
<th>ASSIGNMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>Chapter 1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Getting to Know Your Data</td>
<td>Chapter 2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Data Preprocessing</td>
<td>Chapter 3</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>3</td>
<td>Mining Frequent Patterns, Associations and Correlations</td>
<td>Chapter 6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Exam 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Classification: Basic Concepts</td>
<td>Chapter 8</td>
<td>Team Project 1</td>
</tr>
<tr>
<td>5</td>
<td>Cluster Analysis: Basic Concepts</td>
<td>Chapter 10</td>
<td>Team Project 2</td>
</tr>
<tr>
<td>6</td>
<td>Outlier Detection</td>
<td>Chapter 12</td>
<td>Research Paper</td>
</tr>
<tr>
<td>6</td>
<td>Classification: Advanced Methods</td>
<td>Chapter 9</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Term Paper Presentation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Changes in this course schedule may be necessary and will be announced to the class by the Instructor. The assignments and exams shown are directly related to the Student Learning Outcomes described in Section F.

J. COURSE POLICIES

**Attendance/Tardiness**
You are expected to be in attendance, punctual, and prepared for class. If you are more than 5 minutes late to class, you will be counted as tardy. Please make sure that you will never be tardy to any of your classes or accept the consequences.

**Late Work and Make-up Exams**
NO makeup exams, labs, or quizzes will be allowed unless I have agreed prior to the exam, lab, or quiz time and been provided with official supporting documents.

**Extra Credit**
There is NO EXTRA CREDIT - don't bother asking.
Cell Phone Use
You are required to turn off your cell phone in class and pay attention to class discussions.

Laptop Use
Use of laptops and other electronic devices is restricted to taking notes.

Food in Class
Eating food in class is Not Allowed.

Missed Exam
Missed exams will be graded as ‘0’.

Participation
Class discussions and information provided in class are considered regular course material; it is your responsibility to take appropriate notes. You are expected to attend lectures and actively participate in class discussions.

Others
Read Section L!!!

K. COLLEGE AND UNIVERSITY POLICIES

- **Academic Integrity (University)**
  It is expected that university students will demonstrate a high level of maturity, self-direction, and ability to manage their own affairs. Students are viewed as individuals who possess the qualities of worth, dignity, and the capacity for self-direction in personal behavior.
  See Full University Policy at http://catalog.tamucc.edu/content.php?catoid=10&navoid=313#Academic_Integrity

- **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.

- **Deadline for Dropping a Course with a Grade of W (University)**
  The grade of W will be assigned to any student officially dropping a course. Please
consult with the instructor before you decide to drop to be sure it is the best thing to do. Just stopping attendance and participation WILL NOT automatically result in your being dropped from the class. Should dropping the course be the best course of action, visit the Office of the University Registrar for the Course Drop Form that must submitted. No student is eligible to receive a W without completing the official drop process by this deadline. Please consult the Academic Calendar (http://www.tamucc.edu/academics/calendar/) for the last day to drop a course.

- **Grade Appeals (College of Science and Engineering)**
  As stated in University Procedure 13.02.99.C2.01, Student Grade Appeal Procedures, 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 upon 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, including the responsibilities of the parties involved in the process and the number of days allowed for completing the steps in the process, see University Procedure 13.02.99.C2.01, Student Grade Appeal Procedures. These documents are accessible through the University Rules website at http://www.tamucc.edu/provost/university_rules/index.html, and the College of Science and Engineering Grade Appeals webpage at 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, the Office of the College of Science and Engineering Dean, or the Office of the Provost.

- **Disability Services**
  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 (361) 825-5816 or visit Disability Services in Corpus Christi Hall 116.

  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.

  http://disabilityservices.tamucc.edu/

L. **OTHER INFORMATION**

These guidelines are designed to inform scholars of their responsibilities and of the course requirements in order to make this course a positive experience. The instructor is always available for consultation and discussion with students on any
aspect of a course and of these general guidelines.

1. Consider yourself as a **scholar** rather than a student. The term “student” may imply some passivity, whereas the term “scholar” implies active participation, understanding and searching. We will use these terms interchangeably with the meaning of “scholar” implied. Osmosis does not work in a learning environment!

2. Further, define yourself as a “thinking explorer”. You are responsible for your education; an instructor can only be a guide and a facilitator. An instructor cannot learn for you. If you come across something that really interests you, explore it further.

3. Your experience at this University should not consist of passing a series of courses to earn a degree. Your experience should rather be a series of activities that will give you an education.

4. Concentrate on “learning to learn”. You will have to be a life-long learner to survive in your chosen career.

5. There is no such thing as a stupid question; there is such a thing as a stupid answer. So ask questions, the instructor is taking all the risks! Ask questions of your instructor and of your fellow scholars. Many times questions are more important than answers.

6. The Internet is a tremendous resource and also a great danger. When you find information on the Internet, you have no idea if it is correct. View such information with caution. But, use the Internet to explore topics that interest you. Do not only prepare for the exam in a course – learn as much as you can on the topics introduced to you by the course material. You are responsible for the extent of your education! **READ MINDFULLY !!!!**

7. In addition to details of the syllabus given in class, the syllabus for the course includes all the chapters of the required textbook/s unless indicated otherwise by the instructor.

8. The final letter grade for the class will be based on the raw composite numerical score obtained from the weighted average of the tests, quizzes, exams, labs, etc. as indicated by the instructor. The raw composite numerical score may be adjusted (curved) based on the highest score, the statistical profile of the scores and other academic standards or other considerations. Generally the letter grade of A is 90% and over of the adjusted score, a B is between 80% and 89% (inclusive) of the adjusted score, a C is between 70% and 79% (inclusive) of the adjusted score, a D is below 70% of the adjusted score and an F is below 65% of the adjusted score. An incomplete (I) will only be given in very unusual circumstances. The University regulations on incomplete grades state: “An incomplete notation may be given to a student who is passing but has not completed a term paper, examination, or other required work for reasons beyond the student’s control other than the lack of time”. Students are expected to take ALL tests, quizzes, exams, etc., and to complete and hand in all labs and other assignments. There is no provision for “extra credit”. No final grades will be given via the telephone, e-mail, etc.

9. All University rules, regulations and expected student conduct apply to this course. Students are held responsible for the information given in the current Catalog and Student Handbook.
10. All labs, assignments, etc. must be handed in on the assigned due date. Scholars having problems must notify the instructor well before the due date. Marks will be deducted for poor and sloppily presented work.

11. Labs, etc. handed in after the due date may be subject to a penalty of loss of marks. Labs, etc. handed in after the graded labs, etc. have been returned to students will get zero marks but must be handed in to the instructor.

12. Scholars are asked to take special note of the penalties, which the University attaches to Academic Dishonesty. Consult the Student Handbook.

13. All work handed in to the instructor must be the student's own work. Extracts, excerpts, etc. from the work of others must be suitably noted, acknowledged and properly referenced. Any Group Work will be judged in the same way. That is, it is the work of the group and the extracts, excerpts, etc. of others must be acknowledged.

14. All written and graphical work handed in must be presented neatly printed. Student’s written work will be judged on written communication skills, critical thinking and problem solving ability.

15. There are NO provisions for making up missed exams except in cases where prior arrangements have been made and agreed to by the instructor.

16. Students must keep their given university e-mail address (i.e. firstname.lastname@islander.tamucc.edu). This will be the means of the instructor communicating with students.

17. All work submitted to the instructor (via e-mail or other means) must be clearly marked with the student’s name and the name and number of the course – this is especially important when work is submitted as an attachment to an e-mail.

18. The instructor reserves the right to make changes to the above with due notice to the students. These changes will be announced to the class (see 16 above) and each student is responsible for keeping herself/himself informed of such changes.

GENERAL DISCLAIMER

I reserve the right to modify the information, schedule, assignments, deadlines, and course policies in this syllabus if and when necessary. I will announce such changes in a timely manner during regularly scheduled lecture periods.