Data Mining
CSCI 334
Spring 2013

Professor: Paul Anderson, Ph.D.
Office: 212 J.C. Long
Office Hours: M: 9 – 10 AM
W: 2 – 3 PM
F: 9 – 10 AM
And open door policy

Contact Information: andersonpe2@cofc.edu
Office Phone: 953-8151
Facebook: andersonpe2@cofc.edu
Facebook group: https://www.facebook.com/groups/475661622480402/

Course Description: A course covering data mining concepts, methodologies, and programming. Topics include decision tables and trees, classification and association rules, clustering, pattern analysis, and linear and statistical modeling. Additional topics may include data cleaning and warehousing and techniques for text and web mining.


Course (learning) outcomes:

1. Know the meaning of data mining, some of the application areas and disciplines that use data mining, and understand some of the current major challenges in data mining.
2. Recognize that data mining is part of a larger process and be able to describe the various stages of that process.
3. Understand the need for and techniques for carrying out data cleaning and other data pre-processing activities and to apply them to real-world data sets.
4. Understand and apply a wide range of the fundamental classification and prediction algorithms, including algorithms for decision trees and rule-based classifiers, Bayes classification methods, and other classification approaches such as logistic regression, k-nearest neighbor, and neural networks.
5. Examine and apply metrics for classifier performance and selection.
6. Examine and apply metrics for association pattern evaluation.
7. Understand and apply several clustering algorithms including k-means clustering and BIRCH clustering.
8. Examine and apply metrics for cluster evaluation such as clustering tendency, number of clusters, and clustering quality.
10. Recognize some of the current data mining trends and research frontiers.
11. Explore the use of data mining techniques on different datasets using software packages.

Prerequisite: CSCI 221, MATH 207, MATH 250


Facebook Group (required): https://www.facebook.com/groups/475661622480402/

Course Workspace: https://c9.io/w100pea/csci-334-spring-2013

Grading Policy:
1. Examinations 50%
2. Homework 20%
3. Comprehensive Final 30%
Grading Scale: A: 90-100; B: 80-89; C: 70-79; D: 65-69; F: <65. Plusses and minuses will be used at the discretion of the instructor.

Homework Policy: Homework will be assigned each week and **turned in every Friday**. Homework will consist of a mix of written and programming assignments. The programming assignments will be graded through the cloud9 IDE. You must create your own private workspace. Cheating/sharing code will result in a zero on the assignment and a report to the judicial board. Written homework will placed under my office door.

Exam Policy: Student performance will be assessed through weekly examinations and a comprehensive final exam. Examinations will consist of 2 – 4 questions every Friday. We will immediately go through the solutions after the examination. I will drop the lowest examination score at the end of the semester. If you miss an examination, this will be counted as your drop grade.

Cheating: Students are expected to work independently in this course. Collaborations on specific assignment details are a violation of the honor code. Use of another student's answers is considered cheating, and cases of this nature will be taken to the Judicial Board.

Attendance Policy: Attendance at all lectures is required. Excused absences for illness, personal/family emergency or academic/professional commitments will be granted at the discretion of instructor.

Disability Accommodation: Any student who feels that he or she may need an accommodation due to a disability should speak to me individually to discuss your specific needs. For additional help please contact the College of Charleston Center for Disability services at [http://www.cofc.edu/~cds/](http://www.cofc.edu/~cds/).

Electronics Devices: The use of electronic devices, both stand-alone and network capable, will play an increasingly important roll in teaching and learning at the College of Charleston, including their use in our classrooms. The following policy specifies which electronic devices and network connections can be used and when their use is disallowed in this class.

Devices that are **allowed** to be used at certain times:

**During class, except during tests and exams:**
- Allowed are mobile computing devices, e.g. laptops, palmtops, tablets, electronic pens, calculators. Mute the sound.
- The use of these devices is encouraged for accessing WebCT, taking notes and running simulations during class.

**During tests, exams and quizzes:**
- No electronic devices are allowed to be powered up, unless otherwise specified by the instructor. All books and notes are to be stowed below desk level.

Network Access:
Students may use wired, WiFi and IR networks available during class, whenever electronic devices are allowed, provided the use of the network does not distract other students or the instructor.

Be considerate and sensitive to others. All student behaviors are subject to the policies in the College of Charleston Student Handbook.