Introduction to Discovery Informatics
Discovery Informatics 101
Fall 2012

Professor: Paul Anderson, Ph.D.
Office: 212 J.C. Long
Office Hours:
M: 1 – 2 PM
W: 4 – 5 PM
F: 2 – 3 PM
And open door policy

Contact Information: andersonpe2@cofc.edu
Office Phone: 953-8151
Facebook: andersonpe2@cofc.edu
Facebook group: https://www.facebook.com/groups/156752447781753/

Course Description: Introduction to the use of computer based tools for the analysis of large data sets for the purpose of
knowledge discovery. Students will learn to understand the Discovery Informatics process and the difference between
deductive hypothesis-driven and inductive data-driven modeling. Students will have hands-on experience with various on-
line analytical processing and data mining software and complete a project using real data.

Required Text: None. Readings will be provided online or in printed form for this course.

Required Software: http://www.mathworks.com/academia/student_version/

Course Schedule: http://ds.cs.cofc.edu/index.php/Fall_2012_Introduction_to_Discovery_Informatics

Course (learning) outcomes:
1. To gain an overview the field of knowledge discovery
2. To be able to distinguish and translate between data, information, and knowledge
3. To learn how to store, query, aggregate data in databases
4. To be able to distinguish problems based on computability
5. To learn how to implement distributed computing and storage
6. To apply algorithms for inductive and deductive reasoning
7. To learn introductory and state-of-the-art data mining algorithms
8. To apply data mining, statistical inference, and machine learning algorithms to a variety of datasets, including text,
   image, biological, and health
9. To apply information filtering on real world datasets
10. To apply information validation on real world datasets
11. To apply artificial intelligence concepts to real world datasets
12. To understand the social, ethical, and legal issues of informatics and data science

Course Prerequisite: The course has no prerequisites other than your interest and IT skill fluency. The course is accessible
to all students who have an interest in a breadth-first overview of the field of discovery informatics.

Course Web Page: All materials associated with this course will be posted on the course web page or through OAKS.
http://ds.cs.cofc.edu/index.php/Fall_2012_Introduction_to_Discovery_Informatics

Homework Policy: Homework assignments will be distributed throughout the semester to assist in test preparation and
material synthesis. See schedule for specific dates.

Exam Policy: Student performance will be assessed through two quizzes, one midterm, and one comprehensive final exam,
as indicated on the course schedule.
Grading Policy:
1. Quiz #1 5%
2. Quiz #2 5%
3. Midterm 30%
4. Comprehensive Final 20%
5. Homework 10%
6. Semester Team Project 30%
   a. Idea: Presentation and Write-up 10%
   b. Initial Results: Presentation and Write-up 10%
   c. Final Results: Presentation and Write-up 10%

Semester Team Project: Teams of 2 – 3 people will be created during the first week of class. The first priority of team construction will be to balance those with programming skills with those with domain experience. The project will be graded in three parts. The steps of the project are as follows:
1. Form a team of 2 – 3 people
2. Identify a project from one of the challenges located at Kaggle: http://www.kaggle.com/
3. Present project idea to class and submit 2 page write-up (single spaced) with a minimum of 3 journal article citations
4. Work on the project
5. Present initial results and write-up (3 – 4 page write-up, single spaced, 3 or more journal article citations)
6. Continue work on the project
7. Present final results and write-up (4 – 6 page write-up, single spaced, 3 or more journal article citations)

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.

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/.

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.