Syllabus: CIS 6930LAD Spring 2013

Data Science: Large-scale Advanced Data Analysis

Catalog Description: See UF course catalog.

Pre-requisites and Co-requisites:

Information and Database Systems I (CIS 4301), Data Structures and Algorithms (COP3530), Probabilities and Statistics (STA 5325/5328) or equivalent is a pre-requisite.

Course Objectives:

In this course, we will discuss recent work and publications in Data Science/Big Data Analytics research with emphasis on systems and algorithms for large-scale advanced data analysis. Each student will be responsible for presenting one or more research papers in class and participating in discussions on papers presented by the other students in class. Also, each student will do a class project that has the largest impact on the final grade. Every student should be comfortable with programming (C/JAVA/C++/Scala) and preferably have prior experience with data mining or machine learning.

Instructor:

Daisy Zhe Wang, Office: E456, E-mail: daisyw AT cise DOT ufl DOT edu. Office Hours: Tu/Th 5:00-6:00pm or by appointment.

Course Information:

- **Credits:** 3
- **Section:** 127E
- **Meeting Times:** Tuesday: 8-9th (3:00 to 4:55pm), Thursday: 9th (4:05 to 4:55pm)
- **Where:** CSE Building, E119
- **Teaching Assistant:** N/A
- **Laboratory:** N/A
- **Material and Supply Fees:** None
- **Class web page:** http://www.cise.ufl.edu/class/cis6930sp13lad/

Textbooks and Software Required: None.

Recommended Reading:

We will use papers as our main source. See class page for recommended reading and books.

Course Outline and Topics:

This course will cover the most recent developments in a broad range of Data Science problems. I would like to put more focus on systems and algorithms that enable advanced (statistical/machine learning) data analysis. The topics are as follows:

- Big Data Analysis Systems and Frameworks
  - Map-Reduce – Mahout, Spark/Shark
  - Parallel DB – MADLib, Tuffy/Felix
  - Others: GraphLab, SciDB, DataPath
- Big Data Analysis Models and Algorithms
  - Structured Data Mining
  - Text Analysis
  - Image Retrieval
  - Unsupervised Learning
  - Dimensionality Reduction
- New Research Trends and Applications
  - Crowd-sourcing, Human intelligence
  - Probabilistic Databases, Knowledge Bases
  - Data Visualization, Data Cleaning, Data Integration
  - E-discovery, EMR

Attendance and Expectations:

- I strongly encourage class attendance and participation.
- Please return your reviews, project, and presentations in time. Late returns will cause 20% deduction in your grade for that project/presentation for each late day.
- If I postpone or cancel the office hour, I will post it in the announcements.
- Please avoid any activities that will disturb the flow of the lectures: Silence your cell phones, pagers, etc.

Grading Policy – Methods of Evaluation:

project (55 %)
paper presentations (20 %)
literary reviews (20 %)
class participation (5 %)
novelty in Project (5% bonus)

Grading Scale:

I will not use curve to determine the letter grades. Roughly the boundaries will be:

- 90 -- 100 A
- 85 -- 89 B+
- 80 -- 84 B

- 75 -- 79 C+
- 70 -- 74 C
- 65 -- 69 D+
- 60 -- 64 D
- 0 -- 59 E

Makeup Exam Policy: N/A

Honesty Policy:

All students admitted to the University of Florida have signed a statement of academic honesty committing themselves to be honest in all academic work and understanding that failure to comply with this commitment will result in disciplinary action. This statement is a reminder to uphold your obligation as a UF student and to be honest in all work submitted and exams taken in this course and all others.

Accommodation for Students with Disabilities:

Students requesting classroom accommodation must first register with the Dean of Students Office. That office will provide the student with documentation that he/she must provide to the course instructor when requesting accommodation.

UF Counseling Services:

Resources are available on campus for students having personal problems or lacking clear career and academic goals. The resources include:

- University Counseling Center, 301 Peabody Hall, 3921575, Personal and Career Counseling.
- SHCC mental Health, Student Health Care Center, 3921171, Personal and Counseling.
- Center for Sexual Assault/Abuse Recovery and Education (CARE), Student Health Care Center, 3921161, sexual assault counseling.
- Career Resource Center, Reitz Union, 3921601, career development assistance and counseling.

Software Use:

All faculty, staff and student of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.