Syllabus: CAP4773/CIS6930 Fall 2014

Projects in Data Science

Catalog Description: Projects in data science will facilitate individual projects in data science application areas such as text and image extraction and retrieval, knowledge base construction, pattern mining, social networks, bioinformatics, computational fluid dynamics, vision etc.

Pre-requisites and Co-requisites:

Data Structures and Algorithms (COP3530) or equivalent.

Introduction to Data Science or equivalent.

Optional but preferred: Information and Database Systems I (CIS 4301) and an introductory course to probabilities and statistics are pre-requisites.

Course Objectives:

In order to address the growing need from both industry and academia (e.g., medical and bioinformatics, financial, law enforcement, economics, decision support, social networks) for big data analytics skills including, data management, data mining, machine learning and data visualization, this course is part of the three-course series in the Data Science curriculum. Building on the foundations of databases and data mining, this course will guide students to work in groups on individualized projects in interest areas such as text and image extraction and retrieval, knowledge base construction, pattern mining, bioinformatics, vision and imaging, sensor and social networks, computational neuroscience, medical informatics and scientific data analysis.

Instructor:

Daisy Zhe Wang, Office: E456, E-mail: daisyw@cise.ufl.edu, Office Hours: Mon/Wed 3:50pm-4:50pm or by appointment. Review of Data Manipulation, at Scale and Graph/Text Data Analysis & Communicating Results. Lead group projects.

Sanjay Ranka, Office: E532, Email: ranka@cise.ufl.edu Office Hours: Mon 2-3pm Fri 4-5pm or by appointment. Review Statistical Analytics and Parallel Computing. Lead group projects

Course Information:

- Credits: 3
- Section: 18B9/13GE
- Meeting Times: MWF: 8th (3:00 to 3:50pm)
- Where: CSE Building, E119
- **Teaching Assistant:** Yang Chen and Manu Sethi
- **Laboratory:** N/A
- **Material and Supply Fees:** None
- **Class web page:** TBA

**Textbooks Required:** None

**Software Required:** TBA.

**Recommended Reading:**


b) Various papers in Big Data, Data Science and Data Mining conference (e.g., KDD, ICDE, ICDM)

**Course Outline and Topics:**

The aim is to review and apply data science and big data analytic tools to develop domain-specific applications.

- Review of SQL and No-SQL Technologies
- Review of Data Mining
- Projects in Data Science
  - Problem Definition, Data collection
  - Input and Output Description
  - Model and Method Development
  - Algorithm and System Development
  - Performance Metrics and Visualization
  - Presentations
- Application Areas:
  - Image and Document Retrieval
  - Knowledge Base Construction
  - Bioinformatics
  - Vision and imaging
  - Sensor Networks
  - Computational Neuroscience
  - Natural Language Processing
  - Medical Informatics
  - E-discovery
  - Scientific Data Analysis (e.g., Ecology, Astronomy)
  - Data-Driven Quantitative Education

**Attendance and Expectations:**

- We strongly encourage class attendance and participation.
- Please return your projects 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**

Projects (90 %): proposal (20%), mid-term evaluation (20%), final presentation (25%), final report (25%)
Class Participation (10%)
Project Novelty Bonus (5%)

**Grading Scale:**

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

The boundary for A-, B-, C- will be decided at the end of the semester.

**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.