**INLS 690-163**

**Information Analytics**

**Spring 2015**

Thursday, 9:30 – 10:45am, Manning 214

**Instructor:**

Arcot Rajasekar

**Office:** Manning 021

**Office Hours:** 12:30 – 2:30pm Tue, and by appointment

**Email:** rajasekar at unc dot edu

**Course Description:** The data explosion experienced by computerization of every aspect of our lives from social media to internet of things requires a deeper look at information analytics. The course introduces proven and emerging analytical techniques that can be used to deal with mountains of mostly unstructured data. We will look at several analytical paradigms from Predictive Modeling to Data Mining, Text Analytics to Web Analytics, Statistical Analysis to novel paradigms in Map Reduce and Storm. *Knowledge of programming is essential.*

**Prerequisite(s):** INLS 560 or equivalent

**Textbook** (recommended)**:**  Data Science for Business, Foster Provost and Tom Fawcett

**Grading Scheme:**

1. Class participation 5%
2. Blogs and Journal 10%
3. Homeworks 25%
4. Programming Projects 30%
5. Exams 30%

**1. Course Objectives**:

* Explore the fundamentals of information analytics in areas including statistical analytics, data mining, web analytics, and big data analytics.
* Examine applications of large information analytics
* Gain experience with projects in information analytics.

**2. Hardware and Software Requirements**

We will be using open source software which will require installation and administration. SILS/UNC servers will be used for some of the projects. You may also be required to install and administer some of analytics packages on your laptop for smaller projects and homeworks.

**3. Graded Work**

Your grade will be based on class and blog participation, keeping a journal, a technology paper and presentation, and through projects, homeworks and a final exam, weighted as shown under “Grade Weighting” on the first page.

**Participation**

I require all students to participate actively in class discussions throughout the class. At the beginning of each class, we will have a common discussion period, where we will discuss current events related to topics in the course. I expect that every student reads the ‘required reading’ list, posted at least a week before the class. As the class proceeds, I will be looking for questions, comments and a lively dialogue on the presented material as well as on the required reading materials. Apart from class participation, I also expect students to actively participate in blog posts on topics related to the course. Sometimes I will start a thread of conversation, but I also expect students to take initiative in starting new threads of discussions. The sakai site has facilities for blogs. I have also turned on the chat feature for our course in sakai to enable interactive discussions. There will be no homework – apart from the assigned reading list.

**Journal**

Each student is expected to maintain a journal. This is something of a personal digital library where one will keep all materials related to this course, gathered in the course or elsewhere. I expect material beyond the reading list to be part of your journal. Current events and class discussion topics can also be part of your journal. I also expect tags, metadata and your own commentary added for each material as an outcome of your reading the material. I would strongly recommend the use of the SILS Lifetime Library (<http://lifetime-library.ils.unc.edu/> ) for maintaining the journal as it allows controlled sharing. Please make the material readable by me so that I can evaluate the progress. This journal will be a persistent digital library that may help you later after the course and which you can grow as you gather more relevant material.

**Homework and Project work**

I am planning on a series of home works and projects with Python, R and data analytics tools. More information will be available as the course proceeds.

**Exams:**

**Mid term Exam:**  **February 26, in class**

**Final Exam: FRIDAY, May 1, 8:00AM**

**4. Grading Policies**

The following grade scale will be used AS A GUIDELINE (subject to any curve):

Graduate Percentage Undergraduate Percentage

H 100-95% A 100-90%

P+ 94-90% B 89-80%

P 85-89% C 79-70%

P- 80-84% D 69-60%

L 70-79% F Below 60%

F Below 70%

This scale will be used as a GUIDELINE ONLY. The final grade scale may differ.

**Due Dates and Late work**

Project and paper assignment will have a due date and time and will include instructions for submission. Late

submissions will be given a late penalty. Typically, a late penalty of 10% per day will be applied unless prior arrangements have been made with the instructor.

**Requests for extensions and Absences**

Any request for an extension must be made, preferably by email, at least 24 hours prior to the due date. Written documentation is required for illness. If a serious illness prevents you from taking part, send your instructor an e-mail message, or a friend with a note, describing your condition before schedule. Also, to establish a valid excuse for an illness you must get a note from a physician or the University infirmary.

**Statute of limitations**

Any questions or complaints regarding the grading of an assignment or test must be raised within one week after the score or graded assignment is made available (not when you pick it up).

**5. Course Communication (Sakai)**

Sakai-based course website has been set up and it is the responsibility of every student to **check the Sakai website regularly** for announcements and materials. The Announcements section of the website will be the source for all **official announcements** related to the class. Your instructor may announce tests, assignments, or changes to assignments in class, but there is no guarantee or promise that such announcements will be made in class. The Announcements section of the website is the **only** official, reliable source for announcements, changes, etc. from the instructor. If something the instructor says in class conflicts with information posted by the instructor on the website, then the information posted on by the instructor **on the Sakai website takes precedence**. Verbal instructions are easily misinterpreted, and they do not leave a documentation trail. All students should be able to access the system.

**6. Honor Code**

The UNC Honor Code is in effect for all work in this course. When work or ideas are not your own, you must attribute them. Unless otherwise stated, all assignments in this class are individual assignments, meaning that the substance of the work you turn in must be your own. If you have any doubts or questions about a course of action or a specific situation, please ask for clarification. Students should NOT receive (or give) major creative assistance or ongoing minor support on individual assignments. If you have any questions about this, please ask me.

**7. Special Accommodations**

If any student needs special accommodations, please contact the instructor during the first week of classes.

**8. TentativeTimeline**

|  |  |  |
| --- | --- | --- |
| **Cl No** | **Date** | **Topics in Information Analytics** |
| 1 | Jan 08 | Review of Programming, Databases,  Introduction to Information Analytics |
| 2 | Jan 13 | Introduction to Data Mining |
| 3 | Jan 15 | What is data and Postgres Installation |
| 4 | Jan 20 | Predictive Analytics & Modeling |
| 5 | Jan 22 | R and RStudio |
| 6 | Jan 27 | Predictive Modeling Contd. |
| 7 | Jan 29 | R Challenge |
| 8 | Feb 03 | Supervised Segmentation |
| 9 | Feb 05 | R Challenge |
| 10 | Feb 10 | Supervised Segmentation |
| 11 | Feb 12 | WEKA |
| 12 | Feb 17 | Regression |
| 13 | Feb 19 | WEKA Challenge |
| 14 | Feb 24 | Regression |
| 15 | Feb 26 | **Mid Term Exam (in class)** |
| 16 | Mar 03 | Model Performance |
| 17 | Mar 05 | Rapid Miner |
| 18 | Mar 17 | Similarity and Cluster Analysis |
| 19 | Mar 19 | KNIME |
| 20 | Mar 24 | Similarity and Cluster Analysis |
| 21 | Mar 26 | Orange |
| 22 | Mar 31 | Model Evaluations |
| 23 | Apr 02 | Model Evaluations & Lift |
| 24 | Apr 07 | Text Mining |
| 25 | Apr 09 | Text Mining |
| 26 | Apr 14 | NLTK and Google Analytics |
| 27 | Apr 16 | Data Mining – Apriori Algorithm |
| 28 | Apr 21 | Data Mining – Genetic Algorithms and Neural Nets |
| 29 | Apr 23 | Recap |
| **30** | **May 1** | **Final Exam 8:00AM (FRIDAY)** |