

INLS 625-001

Information Analytics

Spring 2019

Tu-Th 12:30-1:45 Manning 304

Instructor:

Arcot Rajasekar

Office Hours: 11:00 – 12:15 PM Tuesday/Thursday or by email 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: Data Science for Business, Foster Provost and Tom Fawcett

Grading Scheme:

1. Forums	5%
2. Journal	5%
3. Data Lab	30%
4. Exams	30%
5. Project	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 and lab work in information analytics.

2. Hardware and Software Requirements

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

3. Graded Work

Your grade will be based on blog participation, keeping a journal and through projects, homeworks, data lab and two exams, weighted as shown under “Grading Scheme”.

Forum Participation: I require all students to participate actively in the sakai forum throughout the semester. I expect the forum to be the electronic meeting place for students to know each other on the web. I expect every student to read and discuss the classwork and readings that are given for the week in their posts and comment on other students’ postings. Sometimes I may start a thread of conversation, but more often, I expect students to take initiative in starting new threads of discussions.

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. For this purpose we will use the SILS Lifetime Library (<http://lifetime-library.ils.unc.edu/>) as it allows one to keep digital documents, organize them into folders, attach metadata and also perform controlled sharing. Please make the material readable by me (by sharing it with 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. Please look at the web site for more details. Tutorial videos are available there for learning how to use the lifetime library.

Data Lab: We will become familiar with several analytics and data mining tools. Most of these will be open free software which you will be downloading and installing on your laptops. During the course we will have several lab 'sessions' where you will do mini-projects using these tools. You will keep a 'lab notebook' to record details of the lab work. These note books should be maintained in the SILS Lifetime library. The lab schedule is in Section 8.

Project: There will be one term-project. It will be in three phases. The three phases will be worth 10, 15 and 5 points respectively. Details of each phase will be given as we progress. Phase deadlines are shown in Section 8.

Unit Tests: The course is divided into 14 units, the last one being a recap of the course. We will have a series of unit tests, each given after one or more units have been completed. I am planning these tests in lieu of formal exams. Each unit test will be taken on the web using sakai. The schedule for the unit tests are given in Section 8. There will be a 24 hour window for taking these unit tests. Unless noted otherwise, unit tests will be held on Wednesday-Thursday, Noon to Noon timeframe on the week they are scheduled. Of the six unit tests, for each student, the lowest scoring one will be dropped when computing the final grade. Each unit test will be of equal value.

4. Grading Policies:

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

Undergraduate Percentage Graduate Percentage

A/A- 100-94-90% H 100-90%

B+/B/B- 90-87-84-80% P 89-80%

C+/C/C- 80-77-74-70% L 79-70%

D+/D 70-65-60% F Below 70%

F Below 60%

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

Due Dates and Late work: Project, lab and homework assignments will have a due date and time. 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. Schedule

Wk	Tu, Th	Class Topics (Tue)	Lab Work (Thu)	Exams (Units) Wed-Thu	Project Work
1	Jan 10 (Thu)	1.Introduction to Information Analytics & Data Mining	Life-time Library		
2	Jan 15,17	2.Data Types, Terminology & 3. Predictive Modeling	R and R Studio		Introduction to Project
3	Jan 22,24	3. Predictive Modeling	R and R Studio		
4	Jan 29,31	4. Supervised Segmentation	Weka	1,2,3	
5	Feb 05,07	4. Supervised Segmentation 5. Regression	Weka		
6	Feb 12,14	5. Regression	KNIME		Project Proposal Due Feb 16th by midnight
7	Feb 19,21	6. Model Performance & 7. Validation	KNIME	4,5	
8	Feb 26,28	8. Similarity and Cluster Analysis	Rapid Miner/Orange		
9	Mar 05,07	8. Similarity and Cluster Analysis	Google Analytics	6,7	
		SPRING BREAK			
10	Mar 19,21	8. Similarity and Cluster Analysis	Google Analytics		
11	Mar 26,28	9.Model Evaluations & 10. Probabilistic Modeling	NLTK	8	
12	Apr 02,04	11. Text Mining	NLTK		
13	Apr 09,11	11. Text Mining & 12. Data Mining	Data Mining		Project Presentation Video Due Apr 20th by midnight
14	Apr 16,18	12. Data Mining	Project	9,10	
15	Apr 23,25	13. Big Data Analytics 14. Recap	Project		Project Report Due Apr 25th by midnight
	May 3rd FRIDAY	Final Unit Test		11,12,13	
Units	Topics				
1	Introduction to Information Analytics & Data Mining				
2	Data Types, Terminology				
3	Predictive Modeling				
4	Supervised Segmentation				
5	Regression				
6	Model Performance				
7	Validation				
8	Similarity and Cluster Analysis				
9	Model Evaluations				
10	Probabilistic Modeling				
11	Text Mining				
12	Data Mining				
13	Big Data Analytics				
14	ReCap				

