

INLS 625-001
Information Analytics
Fall 2022
Tue/Thu 1230-0145
Manning 208

Instructor:

Arcot Rajasekar

Office Hours: Tue 0200-0330 PM

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” by Foster Provost and Tom Fawcett. Publisher: O'Reilly Media.

Grading Scheme:

1. Sakai Forum Blogs	10%
2. Data Lab	30%
3. Exams	30%
4. Term Project	30%

1. Course Objectives:

- Explore the fundamentals of information analytics in areas including predictive analytics, natural language processing and data mining.
- Examine applications of large information analytics problems.
- 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 blog throughout the semester. I expect the forum to be the electronic meeting place for students to exchange information beyond the class. Finding relevant material and posting them here will be highly appreciated. Sometimes I may start a thread of conversation, but more often, I expect students to take initiative in starting new threads of discussions. **Participation (posting your own items, reading posts by others, and responding to a few posts) every week is mandatory. Missing up to four weeks is allowed.**

Personal Digital Library: Each student is expected to maintain a PDL. This is where one will keep all materials related to this course, gathered in the course or elsewhere. I expect material beyond the reading list, current events and class discussion topics to be part of your PDL. For this purpose we will use the SILS Lifetime Library (LTL) (<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 my id: rajaseka) so that I can evaluate your PDL. This will be a persistent digital library that may help you later after the course (so keep notes and slides) and which you can grow as you gather more relevant material. Tutorial videos are available in the webside for learning how to use the lifetime library. A cheat sheet will also be provided in sakai to

get started. Please follow the cheat sheet (in sakai resources folder) as it has two parts and also tells about naming conventions. **All data lab and project works will be submitted to me through the PDL.**

Data Lab: We will become familiar with several analytics and data mining tools. All of these will be open free software which you will be downloading and installing on your laptops. During the course you will perform ten lab 'sessions' where you will do mini-projects using these tools. A Word or PDF lab report is due through the SILS LTL. **If you have several images and text included, please pull them together into one document.** Code, data and other digital artifacts can be placed in LTL as their own files. Lab schedule is found in Section 8 and can also be found in the Schedule tab in sakai. Labs will be assigned for a week and will be due Wednesday of following week, unless noted otherwise. Since the LTL is a reference for later in your career, I would recommend putting as much information in there that can be helpful when you look at a lab project in a couple of years.

Project: There will be one term-project. It will be in two phases. The phases will be worth 10 and 20 points, respectively. Details will be given as we progress. Phase deadlines are shown in Section 8 and in the sakai Schedule.

Exams: We will have two exams during the semester. The schedule and syllabi for the exams are given in Section 8.

4. Grading Policies:

The following grade scale will be used AS A GUIDELINE:

Undergraduate Percentage	Graduate Percentage
A/A- 100-94-90%	H 100-95%
B+/B/B- 90-87-84-80%	P 94.99-80%
C+/C/C- 80-77-74-70%	L 79.99-70%
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, by email, at least 24 hours prior to the due date. Written documentation is required for illness. If an illness or other circumstances prevents you from taking part, send your instructor an e-mail message, if possible, describing your condition as soon as possible.

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 and often** for announcements and materials. The Announcements section of the website will be the source for all **official announcements** related to the class. **University-provided email should be used for all communications to the instructor.** Other email addresses may not be considered official email.

Normal schedule as posted will be followed. In some circumstances leading to changes, the Announcements section and/or Email of the website will be used to inform changes, etc. from the instructor. If something the instructor says in the class/video 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. Email to make sure to convey any information to the instructor. All students should be able to access sakai and the university email system.

Every week, a 75-minute **zoom office session** will be held by the instructor. All students are encouraged to attend this session and use this as a forum for asking questions and clarifying doubts. We will also use this session for show-and-tell of lab work.

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:

Week	Dates	Lesson: Class Topics	Assigned Lab Work	Lab due by midnight	Project Work
1	Aug 16,18	Introduction to Information Analytics & Data Mining	Lab 1: R and R Studio		Introduction to Project: Start Phase I
2	Aug 23,25	Data Types, Terminology Predictive Modeling	Lab 2: R Challenge	Lab 1: Aug 26	
3	Aug 30, Sep 01	Predictive Modeling	Continue Lab 2	NO LAB DUE	
4	Short wk Sep 08	Sep 06: Wellness Holiday Supervised Segmentation	Lab 3: Weka	Lab 2: Sep 09	
5	Sep 13,15	Supervised Segmentation Regression	Lab 4: Weka Challenge	Lab 3: Sep 16	
6	Sep 20,22	Regression	Lab 5: KNIME	Lab 4: Sep 23	Phase I: Proposal Due Sep 25 th by midnight
7	Sep 27,29	Model Performance Validation	Lab 6: KNIME Challenge	Lab 5: Sep 30	Start Phase II without waiting for Phase I feedback
8	Oct 04	Similarity & Cluster Analysis	Continue Lab 6	NO LAB DUE	
EXAM	Oct 06	Midterm Exam: Lessons for Weeks 1 thru 7 (in class)			
9	Oct 11,13	Similarity & Cluster Analysis	Lab 7: RapidMiner & Orange	Lab 6: Oct 14	
10	Oct 18 Short wk	Similarity & Cluster Analysis Oct 20: Fall Break Holiday	NO LAB	NO LAB DUE	
11	Oct 25,27	Similarity & Cluster Analysis	Lab 8: NLTK-part1	Lab 7: Oct 28	
12	Nov 1,3	Model Evaluations Probabilistic Modeling	Lab 8: NLTK-part2	NO LAB DUE	
13	Nov 8,10	Text Mining	NO LAB	Lab 8: Nov 11	
14	Nov 15,17	Text Mining & Data Mining	NO LAB		Phase II: Video Report Due Nov 20 th by midnight
15	Nov 22 Short wk	Data Mining	Happy Thanksgiving		
16	Nov 29 Short wk	Recap	EOC		
EXAM	Dec 02 FRIDAY	Final Exam (2hrs): Lessons for Weeks 8 thru 15: December 02, FRIDAY 1200 NOON (in class) (Note change in day and time)			