

**INLS 625-001**  
**Information Analytics**  
**Fall 2021**  
Web Online

**Instructor:**

**Arcot Rajasekar**

**Zoom Office Session:** Tue 0545-0700 PM

**Other Office hours:** 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” by Foster Provost and Tom Fawcett. Publisher: O'Reilly Media.

**Grading Scheme:**

1. Sakai Forum Blogs	10%
2. PDL	10%
3. Data Lab	25%
4. Exams	30%
5. Term Project	25%

**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 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. **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 website for learning how to use the lifetime library. A cheat sheet will also be provided in sakai to get started. Please follow the cheatsheet 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.** Hence getting the 10% grade for PDL is guaranteed if you so these submissions.

**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 15 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. **There will be a 24-hour window for taking these exams but each exam will be for 2 hours.**

#### 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:

Lesson or Lab Num	Week of	Lesson: Class Topics	Assigned Lab Work	Lab due by midnight	Project Work
1	Aug 18	Introduction to Information Analytics & Data Mining	LTL, R and R Studio		
2	Aug 23	Data Types, Terminology Predictive Modeling	R Challenge	Lab 1: Aug 27	Introduction to Project: Start Phase I
3	Aug 30	Predictive Modeling	Weka	Lab 2: Sep 03	
4	Sep 06	Supervised Segmentation	Weka Challenge	Lab 3: Sep 10	
5	Sep 13	Supervised Segmentation Regression	KNIME	Lab 4: Sep 17	
6	Sep 20	Regression	KNIME Challenge	Lab 5: Sep 24	Phase I: Proposal Due Sep 25 <sup>th</sup> by midnight
7	Sep 27	Model Performance Validation	RapidMiner & Orange	Lab 6: Oct 01	Start Phase II without waiting for Phase I feedback
	Oct 04	<b>NO CLASS, NO LAB</b>			
8	Oct 11	Similarity and Cluster Analysis	NO LAB 8	Lab 7: Oct 15	
	<b>Oct 14</b>	<b>Midterm Exam (2hr): Lessons 1 thru 7: Thursday: 24-hour window</b>			
9	Oct 18	Similarity and Cluster Analysis	NO LAB 9		
10	Oct 25	Similarity and Cluster Analysis	NLTK-part1		
11	Nov 1	Model Evaluations Probabilistic Modeling	NLTK-part2		
12	Nov 8	Text Mining	NO LAB	Lab 10 & 11: Nov 12	
13	Nov 15	Text Mining & Data Mining	NO LAB		
14	Nov 22	Data Mining	Short week		
15	Nov 29	<b>Recap</b>	Short week		Phase II: Video Report Due Dec 1 <sup>st</sup> by midnight
	<b>Dec 07</b>	<b>Final Exam (2hr): Lessons 8 thru 14: Tuesday: 24-hour window</b>			