INLS 625-02W
Information Analytics
Fall 2020

Web Online

Instructor:

Arcot Rajasekar
Zoom Office Session: Wed 0600-0715 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 structured and unstructured data. We will look at several analytical paradigms from Statistical Analytics to Predictive Modeling from Text Mining to Data Mining and Deep Learning. *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	15%
2.	PDL	5%
3.	Data Lab	25%
4.	Exams	30%
5.	Project	25%

1. Course Objectives:

- Explore the fundamentals of information analytics in areas including predictive analytics, natural language processing, data mining and deep learning.
- 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 participation in sakai forum blogs, keeping a personal digital library (PDL) and through projects, data lab work 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.**

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 (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. **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 Lifetime Library. Lab schedule is in Section 8. **Each lab report is due by midnight Monday of the week after in which it was assigned.**

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

Exams: The course is divided into 12 units. 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, midnight to midnight.**

4. Grading Policies:

The following grade scale will be used AS A GUIDELINE:

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 and lab assignments will have a due date and time. Late submissions will incur 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. If an unavoidable circumstance or serious illness prevents you from taking part in any activity, send your instructor an e-mail message describing your condition **before** the scheduled activity. 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 posted.

5. Course Communication (Sakai, Email and Zoom)

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. The instructor may also provide additional subject information during this session.

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 of	Unit Number. Class Topics	Lab Work	Exams(units)	Project Work
	Aug 10	1.Introduction to Information Analytics & Data Mining	LTL, R and R Studio		
	Aug 17	2.Data Types, Terminology 3. Predictive Modeling	R Challenge		Introduction to Project
	Aug 24	3. Predictive Modeling	Weka		
	Aug 31	4. Supervised Segmentation	Weka Challenge		
	Sep 07	4. Supervised Segmentation 5. Regression	KNIME		
	Sep 14	5. Regression	KNIME Challenge		Project Proposal Due Sep 18th by midnight
	Sep 21	6. Model Performance 7. Validation	RapidMiner/ Orange		
	Sep 28	8. Similarity and Cluster Analysis			
	Oct 01	**Midterm Exam (2hr)**	**24-hour window**	**Units 1 thru 7**	
	Oct 05	8. Similarity and Cluster Analysis	Data Mining		
	Oct 12	8. Similarity and Cluster Analysis	Deep Learning		
	Oct 19	9. Model Evaluations 10. Probabilistic Modeling			
	Oct 26	11. Text Mining	NLTK-part1		
	Nov 02	11. Text Mining & 12. Data Mining	NLTK-part2		
	Nov 09	12. Data Mining			Project Video Report Due Nov 12th by midnight
	Nov 16	**Recap**			
	Nov 19	**Final Exam (2hr)**	**24-hour window**	**Units 8 thru 12**	

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