



Benjamin Korallus Arnold
bkorallus@luc.edu
16 East Pearson, Chicago, IL 60611
Office Hours: 30 minutes post lecture

INFS494, Data Mining, Summer, 2018
Schreiber Center 725: Weekdays 9:00am-12:00pm

Course Objectives

The goal of the course is to introduce students to the current theories, practices, tools and techniques in data mining. Specifically, at the end of the course students will be able to:

- Explain how businesses can gain competitive advantage through the mining of data
- Describe when and how various data mining techniques should be applied
- Understand the basic process and mechanics of data mining
- Be able to interpret the results of various data mining techniques and make appropriate strategic recommendations
- Execute basic data mining techniques using R statistical software

Because many topics and concepts in data mining are learned most efficiently through hands-on work with data sets, we will spend time in class analyzing and mining data. As such, personal laptops with R statistical packages installed will be required for class participation.

The goal is to gain a better understanding of how data mining is applied and what is involved in data mining projects.

Textbook (Recommended)

- [ISLR] An Introduction to Statistical Learning, with Applications in R, Gerath James, Daniella Witten, Trevor Hastie and Robert Tibshirani, Springer, ISBN 978-1-4614-7137-0 ISBN 978-1-4614-7138-7 (eBook).
- Available as **free** pdf: <http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf>

Software (Required)

- All homework and in class workshops will be performed in R
- R Studio is the recommended development software. The open source desktop version is available for a free download at: <https://www.rstudio.com/products/rstudio/download/>

Grading

Homework	300
In-Class Workshops and Participation	200
Group Project Presentation	250
Final Exam	250
Total:	1000

Homework

There will be 5 individual homework assignments worth 60 points each. All homework is due at midnight on the due date (the night prior) and should be submitted using the assignment links given in Sakai. All assignments will be expected as an R-markdown file knitted as either an HTML or PDF document submitted via Sakai. Directions and guidance for using the R statistical language and R-markdown will be discussed on the first days of class but will require individual efforts outside of the classroom, especially for those students with limited background.

Homework is individual.

Assignments build upon the previous and thus late submissions will only be accepted for half credit up until (but not accepted after) the due date of the final assignment submission.

In-class workshops, quizzes and class participation

There will be various discussion topics, hands on labs and group project workshops in classes 1 through 8. As a result, attendance and active participation is required and worth 25 points a class. If you need to miss a lecture for a strenuous circumstance, please coordinate with me via email.

Group Project Presentation

The project will serve as a practical learning experience in understanding various issues in data mining. Students should be able to describe their data mining process, the results, and the potential benefits of expected findings.

You are free to choose your own dataset, but it must have at least 2000 rows. Your analysis should address (at least) each of the following issues:

- Why you are looking at this data set?
- What question is driving the analysis?
- How you manipulated the original data?
- What techniques you used and why (how do they relate to your questions)?
 - Groups must use at least 1 unsupervised technique and 1 supervised technique
- How well did the data mining techniques perform on your data set?
- Recommendations for action
- Conclusions

There are two deliverables for the group project: a presentation to the class and an R-markdown report in either HTML or PDF format. These should address the major sections of the CRISP-DM methodology, describe your analysis, code and answer the questions above.

The points for this project/presentation will be allocated as follows:

- 50 points: average of points assigned to you by your other group members
- 100 points: data mining approach and analysis
- 50 points: your part of the presentation and participation on presentation day
- 50 points: quality and completeness of the written documentation (markdown & ppt)

Final Exam

A closed book, final exam will be held on the last class and include both multiple choice and essay questions. The exam will focus on topics from the lecture, not R-programming.

Course Schedule (Tentative)

Class	Date	Topic	Reading	HW Due
1	07/02	Introduction to Data Mining and R Programming	ISLR 1	--
2	07/03	Literate Programming Market Basket Analysis	ISLR 2	swirl() assignments
	07/04	Independence Day - Holiday	--	--
3	07/05	Regression Techniques Linear Regression	ISLR 3	HW-1
4	07/06	Model Evaluation Model Validation	ISLR 5	HW-2
5	07/09	Classification Techniques Logistic Regression	ISLR 4	HW-3
6	07/10	Decision Trees Ensemble Methods	ISLR 8	HW-4
7	07/11	Clustering Techniques K-Means	ISLR 10	HW-5
8	07/12	Additional Methods SVMs, Neural Networks	ISLR 9	--
9	07/13	Project Presentations	--	Group Report
10	07/16	Final Exam	--	--

Grading Scale

A	≥	920
A-	≥	900
B+	≥	870
B	≥	820
B-	≥	800
C+	≥	770
C	≥	720
C-	≥	700
D+	≥	670
D	≥	620
D-	≥	600
F	≤	600

Quinlan School of Business Policies:

Make-Up Examinations

Loyola University academic policy provides that tests or examinations may be given during the semester or summer sessions as often as deemed advisable by the instructor. Because Quinlan faculty believes examinations represent a critical component of student learning, required examinations should be taken during the regularly scheduled class period.

Make-up examinations are discouraged. Exceptions may be granted only by the faculty member or department chair, and only for unavoidable circumstances (illness verified by a signed physician's note, participation in intercollegiate athletic events, subpoenas, jury duty, military service, bereavement, or religious observance). A make-up final examination may be scheduled only with the permission of the appropriate Quinlan Assistant or Associate Dean.

For a student with a documented special testing need, please consult University policy concerning use of the testing center in Sullivan Center at Lake Shore Campus.

Academic Integrity

All members of the Quinlan School shall refrain from academic dishonesty and misconduct in all forms, including plagiarism, cheating, misrepresentation, fabrication, and falsehood. Plagiarism or cheating on the part of the student in individual or group academic work or in examination behavior will result minimally in the instructor assigning the grade of "F" for the assignment or examination. In addition, all instances of academic dishonesty must be reported to the chairperson of the department involved.

For further information about expectations for academic integrity and sanctions for violations, consult the complete Quinlan School of Business Honor Code and Statement of Academic Integrity on the Quinlan website:

<http://www.luc.edu/media/lucedu/quinlanschoolofbusiness/pdfs/Honor-Code-Quinlan-July2012.pdf>

Notes

- This class may occasionally deviate from the stated course outline above. The instructor reserves the right to make changes as needed to the course syllabus. Any issue that is not described in this syllabus will be resolved at the sole discretion of the instructor.
- R is a well-documented language with numerous free online resources. As such, I will not respond to questions or help troubleshooting in R programming via email. I will, however, spend time during office hours and class workshops to address any R programming issues.