
Statistics 523: Applied Data Mining Course Syllabus

Department of Mathematical Sciences, Central Connecticut State University

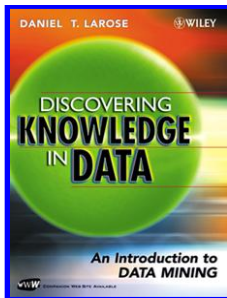
Daniel T. Larose, Ph.D.
Professor of Statistics
Director, Data Mining @CCSU
Office: Marcus White 118
Phone: (860) 832-2862

larosed@ccsu.edu
www.math.ccsu.edu/larose
www.ccsu.edu/datamining
www.dataminingconsultant.com

Course Description

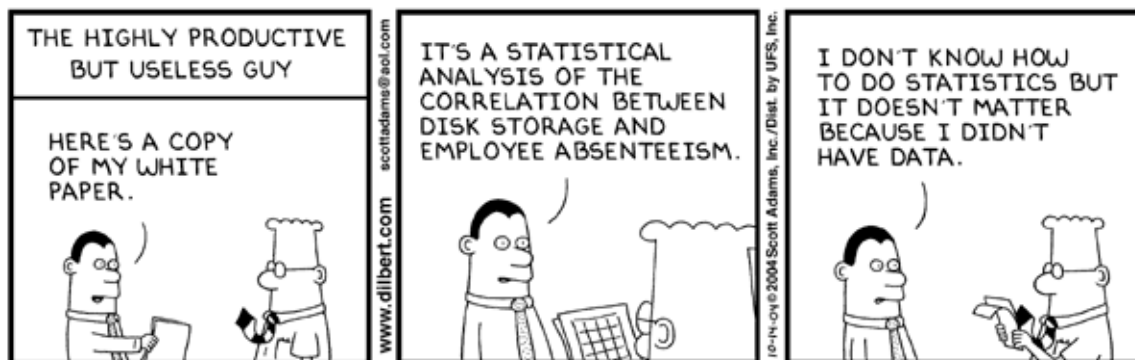
Advanced investigation of data mining models and methodologies. Topics may include dimension reduction methods, Kohonen networks clustering, association rules using the "a priori" and generalized rule induction algorithms, naive Bayes classification and Bayesian networks, and genetic algorithms. 4 credits.

Texts



*Discovering Knowledge in Data:
An Introduction to Data Mining*
By Daniel T. Larose
John Wiley and Sons, Inc., 2005

Data Mining Methods and Models
By Daniel T. Larose
John Wiley and Sons, Inc., 2006



© UFS, Inc.

Course Syllabus
Stat 523: Applied Data Mining
Daniel T. Larose, Ph.D.

Required Software

PASW Grad Pack, latest version. No need to buy the one with text mining for this course.

Data Sets

You may download the data sets used in the books from www.dataminingconsultant.com.

Remote Access to PASW and Other Software

As CCSU students, you may remotely access PASW, SPSS, Minitab, and other software, as long as you are registered for courses at CCSU. The software is located on a Citrix server at CCSU, and may be accessed at <http://citrix.ccsu.edu>. The first time you use it, you will have to download the client application to your machine. In addition to statistical software, there is lots of other software there you may access remotely. You will need to make sure your CCSU student account is set up properly. Go to the secure site at <https://accounts.ccsu.edu>. Questions about the Citrix server (though not about the various applications) may be addressed to the CCSU ITS help desk. Please note that the remote access to PASW is no substitute for purchasing the software, since the remote access is rather clunky for very large data sets like the ones you will be analyzing for the course projects. But it is workable if you would like to play with it with small data sets.



Office Hours

My office hours for the semester are listed on my faculty website at www.math.ccsu.edu/larose. Students living within driving distance of CCSU should feel free to visit for questions and general chit-chat. Other students should feel free to phone during these times. All students should email me to set up an appointment, to make sure that I am not busy with other students or meetings. I am of course a web addict, and usually respond to email within a reasonably short time (depending how difficult your question is!). It is recommended that you network with your fellow students, make friends, and help each other with the homework. The best way to get started networking is to participate enthusiastically in the introductory threaded discussion, where everyone introduces themselves and tells us all about themselves. Very often, we find that "it is a small world", and that we have much in common with our fellow classmates.

Estimated Course Evaluation Plan (Subject to Change)

	Not sure yet.	
We	will probably have two major	Projects.
And	Perhaps some other stuff too.	
More	News	Soon.

Late Submissions = Bad

You will be provided with due dates for all assignments. To avoid last minute computer glitches etc you should get your work in as early as you can. *Late submissions count as zero.* Consider that the due date is a client-driven specification, and your company will lose the contract if the deliverables (report, predictions, etc.) are not submitted by the due date. As the consultant for this project, losing that contract will cost you your job. In other words, if your report is late, you will flunk the course. So get started now.

How Much Time Should I Spend Studying and Working on the Course?

Never for a moment lose sight of the fact that this is a challenging four-credit graduate-level course, and will require your best effort, especially in an online environment. Four-credit graduate courses typically require about 275 hours of student work altogether.

Email forwarding.

WebCT allows you to forward your internal email to another email address. Why not set it up so that course messages get forwarded to an account you use more often?



Copyright © 2001 United Feature Syndicate, Inc.

Withdrawal:

Withdrawals after October 25 are not automatically approved. If a student remains in the course past this date, it represents a commitment by the student to finish the course. This commitment can be broken only if circumstances beyond the control of the student arise. There will need to be extreme extenuating circumstances. So please decide before this date whether you are staying with us for the duration.

Caveat

You are responsible for all material assigned during the semester, the content of the texts and units covered. (Reading the powerpoints is not required, since they simply recapitulate what is in the books. However, some students seem to like them. Reading the books is required.)

Submitting the PASW Project

When you open the PASW Project, please note carefully the instructions on how to submit your work. You will be asked to submit it via email. In addition, you will be required to provide me with a hard copy. US Mail (first class or priority) or the equivalent (Fed Ex, etc) is acceptable. Faxes are NOT acceptable. Mail your material to:

Dr. Daniel T. Larose
Re: Stat 523 Project 1 (or 2 etc)
46 Willow Creek Drive
Tolland, CT, 06084

The PASW project MUST be sent via internal WebCT email. Otherwise, my faculty email box will be overwhelmed, and no-one will be able to contact me.

Respect and Civility

Please remember, the watchword throughout the course is respect, for each other, and for me. After all, we are all in this together.



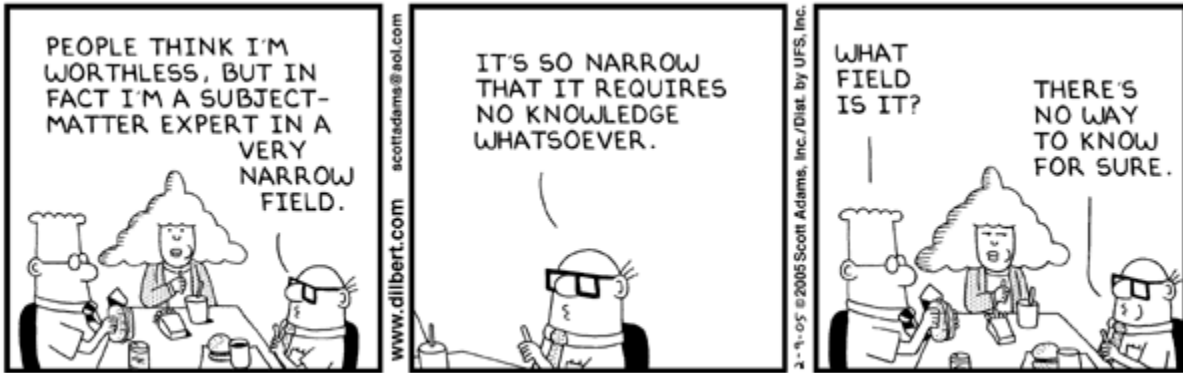
Academic Misconduct

Academic misconduct, including cheating and plagiarism, is prohibited. For more on CCSU's Academic Misconduct Policy, see www.ccsu.edu/AcademicIntegrity.

So, Good Luck in Applied Data Mining

I hope you enjoy it from your side of the monitor as much as I do from mine!
You are on your way to a new Master of Science in Data Mining!

Course Syllabus
Stat 523: Applied Data Mining
Daniel T. Larose, Ph.D.



© UFS, Inc.



© UFS, Inc.