

# CS 419 Informatics

This course presents students with basic, yet seminal techniques for machine learning and data mining. At the end of the course the student should be able to understand how these, and other more sophisticated techniques work and when it is appropriate to apply them. This is a valuable skill in a job market where information management and data mining are key. This is a project based class.

**Meetings:** Tuesdays from 7:05pm – 9:45pm

**At:** BBH 131

**Office Hrs:** Tue-Wed-Thu from 2:30pm – 4:30pm

**Email:** f-iacobelli@neu.edu

**Website:** <http://www.fid.cl/courses/cs419>

**Textbook:** Machine Learning by Tom Mitchell, 1997.

**Syllabus:** Get this document as a PDF [here](#).

Week	To be covered	Assignment
8/30	<b>Administrivia/Introduction:</b> Machine Learning and Data Mining	Opinion Mining with Movie Reviews. Use the reviews in here and turn in your program and a paragraph or two describing your training and test sets, your reasoning and your evaluation.
9/06	<b>General to Specific Learning:</b> Ch. 2	<ul style="list-style-type: none"><li>• Slides</li><li>• Assignment 2: Exercise 2.5 (a) and (b) in the book</li></ul>
9/13	<b>Decision Trees:</b> Ch. 3	<ul style="list-style-type: none"><li>• Slides</li><li>• Assignment 3: Exercise 3.4 (a), (b) and (c)</li></ul>
9/20	<b>Neural Networks:</b> Ch. 4	<ul style="list-style-type: none"><li>• Slides</li><li>• Download Weka and play with it (not graded, but needed for the next assignment)</li></ul>

Week	To be covered	Assignment
9/27	<b>Basic Statistics:</b> Ch. 5	<ul style="list-style-type: none"> <li>• Slides</li> <li>• <b>Homework:</b> download this file (txt_sentoken.arff) and load it in Weka. This file is the movie review data adapted to work with Weka.</li> <li>• Using Weka, evaluate the performance of Neural Networks (with a couple of different number of hidden nodes), Decision Trees (with and without pruning), Naïve Bayes, and one other method, not covered in class (do basic research to know how it works, roughly).</li> <li>• Repeat the last step about 10 times, obtain measures of accuracy, F-measure, precision and recall and compare which one performs the best.</li> <li>• Write one or two paragraph explaining why is it that one method works better than another. What makes it better suited for this problem.</li> </ul>
10/4	<b>Bayesian Learning I</b> Ch. 6; 6.1 – 6.10.	Slides
10/11	<b>Bayesian Networks</b> Ch. 6; 6.11 – 6.13 and attend the talk by professor Neapolitan in LWH 3044 at 7pm.	
10/18	<b>Bayesian Networks</b> Ch. 6; 6.11 – 6.13	
10/25	<b>K-Nearest Neighbor</b> Ch. 8	<ul style="list-style-type: none"> <li>• Compute, using the Naïve Bayes approach, the probability of playing tennis given <math>\langle \text{Overcast}, \text{Cool}, \text{Low}, \text{Weak} \rangle</math>. Use Table 3.2 (p.59) in the book for training.</li> <li>• In the slides for Bayesian Learning above, Slide 43 contains a Bayesian Network. Based on that network, answer the question: if John and Mary call, has there been a burglary? how probable is that? In other words, compute <math>P(\text{Burglary}   \text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true})</math>.</li> <li>• Choose Teams. I have many corpora available</li> </ul>
11/01	<b>NLP For Text Mining</b> Papers TBD	Set up team website and choose a project
11/08	<b>A-Priori Associations</b> Reading to be assigned	Read and Blog about assigned Papers
11/15	<b>Latent Dirichlet Allocation</b>	
11/15	<b>Discussion: Hard Problems in Data Mining</b>	
11/29	<b>Data Mining From the Web</b>	
11/06	<b>Advanced Topics I</b> TBD	
12/13	<b>Final Presentations</b>	

## Assignments

Please turn in your assignments via email. The emails should be addressed to me (email above) and the subject should say cs419-HWx where x is the assignment number.

## **The fine print**

This syllabus is a living document. It is subject to change. If you want to look at the syllabus always download the latest copy from the website.

### ***Every Class***

Come prepared, google the topics, read the book and extra readings and, of course, do the homework on your own.

After choosing your teams, you are required to blog your reactions to the readings and the topics EVERY WEEK. You are also required to comment on, at least one other post.

### ***Grading***

All assignments are worth the same. The project is worth 55% of your grade. Assignments are worth 45%. All of your blogs count as one assignment and every blog missing counts as a letter point discount.

### ***Late Work Policy***

Each day that your work is late the maximum possible grade decreases by one letter point.

### ***Academic Integrity***

Students are required to abide by Northeastern Illinois University's academic integrity policy. Failure to adhere to this policy will likely result in a failing grade in the class and / or expulsion from the University.