

Statistic
COURSE NUMBER: 22:960:641
COURSE TITLE: Analytics for Business Intelligence

COURSE DESCRIPTION

This course is intended for business students of data mining techniques with these goals: 1) To provide the key methods of classification, prediction, reduction, and exploration that are at the heart of data mining; 2) To provide business decision-making context for these methods; 3) Using real business cases, to illustrate the application and interpretation of these methods.

The course will cover Classification (e.g. helps banks to determine who will default on a loan, or email filters to determine which emails are spam), Clustering (like classification, but groups are not predefined, as in legitimate vs. spam email, so the algorithm will try to group similar email together for instance), Regression (e.g. how ad campaigns in offline media such as print, audio and TV affect online interest in the advertiser's brand), Association Rule Learning (enables merchants, for example Amazon, to determine which items customers tend to buy together and make suggestions for further purchase, otherwise known as “market basket analysis”); and Neural Nets (has helped financial agents to model complex currency market trading).

The pedagogical style will use business cases so the student can follow along and implement the algorithms on his or her own with a very low learning hurdle. In addition, students may work in teams to mine their own data. Individual students may request to work on their own company data.

COURSE MATERIALS

Data Mining for Business Intelligence – Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner, G. Shmueli, N. R. Patel & P. C. Bruce (2nd edition)

ISBN-10: **0470526823**

ISBN-13: **978-0470526828**

ERRATA: <http://www.dataminingbook.com/errata>

SOFTWARE

MS Excel and *XLMiner* add-in (included with text) (required) <http://www.resample.com/xlminer/>

PREREQUISITES: Students are assumed to have mastered basic business mathematics and have some good familiarity with EXCEL. An elementary statistics course is assumed.

FINAL GRADE ASSIGNMENT

Homework	60%
Midterm	20%
Project	20%

Statistic (22:960:641)

COURSE SCHEDULE

Date	Reading	Topics
Week 1	Jan 22	Course administration;
		Ch 1,2
		Introduction and overview of data mining process;
		Software: Excel + XLMiner
Week 2	Jan 29	Ch 3
		Data visualization
		Ch 4
		Dimension reduction
Week 3	Feb 5	Ch 5
		Classification methods and evaluation metrics
Week 4	Feb 12	Ch 6
		Multiple linear regression models (MLR)
Week 5	Feb 19	Ch 7
		K-nearest neighbors
		Ch 8
		Naïve Bayes
Week 6	Feb 26	Ch 9
		Classification and regression trees (CART)
Week 7	March 5	
		Midterm exam
Week 8	March 12	Ch 10
		Logistic regression + Topics of projects
Week 9	March 26	Ch 11
		Neural nets
Week 10	April 2	Ch 12
		Discriminant analysis
Week 11	April 9	Ch 14
		Cluster analysis
Week 12	April 16	Ch 13
		Association rules

Week 13	April 23	Ch 15-17	Time series
Week 14	April 30		Project Presentations