LOGOM 6345 Business Analytics and Data Mining  
Fall 2013 - Professor L. Douglas Smith  
(revised 9/10/13)

Office: 220CCB (Computer Center Building)  
Office Phone: 516-6108  
E-mail: ldsmith@umsl.edu  

Home Phone:  (314) 261-5014 (STL)  
Mobile:  (314) 556-7742  
Seabeck, WA:  (206) 493-2915

I am generally in my office and available by appointment during the day M-F. You may drop in at any time. Do not hesitate to call me at home, even late in the evening. I am happy to discuss concepts and to offer help over the telephone. You may also contact me via e-mail. Even when traveling, I check e-mail frequently where practicable.

Overview of the Course

This course concentrates on converting elemental data from various sources into business intelligence. From a series of cases and analytical exercises with large datasets from different business contexts, students gain experience in applying the principles and techniques for business analytics and data mining. The primary focus in the course is on predictive modeling for managerial applications. Within broad frameworks for business analytics and data mining, we employ a rigorous three-step process for constructing, testing and validating statistical models. We experiment with techniques for forecasting qualitative and quantitative outcomes, uncovering patterns in data, clustering entities in groups, and visualizing relationships. Finally, we discuss applications of mathematical optimization and discrete-event simulation for data-based decision making.

Students perform their analyses with professional versions of commercial or open-access software (SAS, including SAS Enterprise Miner, SPSS, IBM SPSS Modeler, Tableau, KNIME, R, etc.) and work in study groups to produce presentations for class discussion.

Student Registration for SAS OnDemand:

In addition to software on the UMSL student network (SAS, SPSS, EXCEL) and open-access software (KNIME, R), we shall use the SAS Enterprise Guide/SAS Enterprise Miner bundle that is available online from SAS® OnDemand and as embedded in the UMSL academic license for SAS9.3. Data for several cases and some current modules for SAS Enterprise Miner will be available only from the SAS® OnDemand website.

Students register for SAS® OnDemand for Academics and then access SAS Enterprise Guide/SAS Enterprise Miner bundle by:

1. Registering at the Web site: http://support.sas.com/ondemand/index.html#account
2. Reviewing the information and following stipulated steps at the Website.

For questions about using SAS® OnDemand for Academics, refer to http://support.sas.com/ondemand

Students are encouraged, in parallel, to complete the e-learning modules for SAS® Enterprise Miner (which correspond to the Text of Applied Analytics Using SAS® Enterprise Miner) and receive a certificate to that effect from SAS Institute.
Course Schedule:

Classes will be held regularly on Thursdays at 7-10 PM and Saturdays at 10AM-2PM on: Sept 12,14; Oct 3,5; Oct 24,26; Nov 14,16; Dec 5,7; and Dec 12, 14. Times may be modified, if agreed unanimously, to accommodate scheduling conflicts that arise and provide sufficient face-to-face time for class discussion.

Course Prerequisites:

LOGOM5300 and INFSYS 5800 or equivalents.

Text and References:

**Applied Analytics Using SAS® Enterprise Miner** will be available to each student as an e-learning supplement after registering for SAS®OnDemand. Class lecture notes on each topic will be provided by the instructor on Mygateway in advance of each class. Other referenced textual material and assigned readings will be available online at designated websites or through ABI/INFORM. (Please keep current with updates to the reading list posted on Mygateway throughout the semester).

**Multivariate Data Analysis** by J.F. Hair et al. (Prentice Hall various editions since 1998) provides good background on statistical techniques for business analytics and data mining.

Approximate Grading Scheme:

Exams (two take-home exams with cases to analyze individually) 60%; Case exercises done by study teams for class discussion 20%; Major project and report 20%. Grade breaks for the final weighted scores usually occur about 80 percent for B+/A-, 70 percent for C+/B-, and 60% for D+/C-.

Course Topics

1. Introduction to BI/BA/DM
   a. Major applications of business analytics
      i. Risk assessment
      ii. Target marketing
      iii. Customer relations management
   b. General frameworks (processes) for business analytics and data mining
      • KDD (Knowledge Discovery in Data), CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, Assess)
2. Analysis involving targeted variables and nonlinear relationships
   a. Application of regression modeling for predicting continuous measures
      i. Accommodating nonlinear relationships with continuous and discrete shifts in the expected values of target (response) variables
   b. Application of logistic regression for predicting categorical outcomes
      i. Bivariate
      ii. Multivariate (multinomial)
   c. Application of Probit models
d. Application of Decision trees
e. Application of Neural networks

3. Discussion of challenges with data integration
   a. Changes in business entities, business environments and managerial practice
   b. Matching data from different sources
   c. Geographical attribution
   d. Time-interval attribution
   e. Missing data
   f. Errors and outliers

4. Building predictive models from data mining
   a. Partitioning data for model construction, testing and validation
   b. Dealing with information redundancy
   c. Transforming data for statistical modeling
      i. Allowing nonlinear forms
      ii. Adjustments for missing data
   d. Avoiding over-fitting
   e. Measuring fit and predictive performance

5. Application of the SEMMA Process in data-mining case studies and class projects
   a. Articulation of project objectives and model-selection criterion
   b. Data integration and cleaning
      i. Checks for reasonable values and consistency
   c. Model development
   d. Model testing
   e. Triangulation and comparison of results from alternative modeling approaches

6. Clustering methods
   a. Clustering objects
   b. Clustering variables

7. Visualizing relationships
   a. Geographical distributions
   b. Frequency distributions
   c. Clusters of objects
   d. Illustrating changes of measures and relationships through time

8. Text mining
   a. Applications for market research and CRM
   b. Illustrations with SPSS Miner, SAS Enterprise Miner

9. Data-based decision support
   a. Illustrations of applied optimizing methods
   b. Illustrations of discrete-event simulation.