ADVANCED BUSINESS ANALYTICS
Winter Term, 2013

Professor Stephen Powell   Assistant**:** Brenda Gray
Buchanan 111
V: 646-2844   brenda.gray@tuck.dartmouth.edu
stephen.g.powell@tuck.dartmouth.edu

Objectives
Business analytics is a set of data analysis and modeling techniques for understanding business situations and improving business decisions. These techniques range from everyday methods such as Pivot Table to advanced methods such as neural networks. Business analytics is conventionally divided into three domains:

- Descriptive – what *is* happening now?
- Predictive – what *will* happen in the future?
- Prescriptive – what *should* happen?

Descriptive methods
Descriptive methods involve using data to describe the current or recent past situation for an organization. For example, one might use such methods to ask how profits are distributed geographically, or which basketball player contributed the most to a winning season. Descriptive methods are usually not closely tied to specific decisions, and involve little to no modeling.

Predictive methods
Predictive methods also rely heavily on data, although some modeling is usually involved. Here the focus is on forecasting future outcomes, normally under the assumption that the driving forces in play in the past will continue into the future. Because of this assumption, predictive methods rely more on data analysis than modeling.

Prescriptive methods
Prescriptive methods answer questions related to what decision makers want to happen in the future. Thus they are most closely tied to the decision making process. Data plays a role in these methods but modeling is the fundamental tool here. Optimization and simulation are prescriptive tools.

This course builds on the core courses in Statistics and Decision Science. It rounds out the student's background in data analysis by adding to the classical statistical tools taught in the core tools from artificial intelligence, machine learning, and data exploration. It develops the student's background in decision science by adding tools ranging from data visualization to time series analysis.

Analytics is associated in many people's minds with the marketing function, presumably because applications in that area have received extensive publicity. While marketing provides many good applications, both operations and finance are increasingly fertile areas for application of analytics. Analytics plays an increasingly important role in sports management, as anyone who has read *Moneyball* knows. And analytic skills are in high demand in the nonprofit and governmental arenas. In fact, analytics is even a mission-critical skill in military, intelligence and security operations.

While examples related to marketing will occasionally be used in this course, the majority of the applications relate to finance, operations, sports, medicine, or other domains. Here are a few of the questions addressed by these methods:

- Can we identify which banks are most likely to default?
- Can we predict which flights from a given airport are most likely to be delayed?
- Can we determine which data are most useful for allowing us to identify web users most likely to respond to an offer?
- Can we use machine learning techniques to develop a method for identifying songs that will appeal to web radio listeners?
- Can we accurately forecast the demand for public transportation?
- Can we create metrics that will accurately capture the contribution of an individual player to a sports team?

Requirements

Class Preparation and Homework
Preparation for class will typically consist of watching a video lecture and a software demonstration. Homework will involve analyzing a business problem using the technique described in the lecture, and submitting results electronically. All classes will include homework.

Project
Students will complete a project on a topic of their choosing. At a minimum, a project will involve
1. Identifying a question to answer
2. Locating appropriate data
3. Using one or more analytic methods to address the question
4. Presenting results

Office hours
I will hold normal office hours on Tuesdays from 2-4:00 in Buchanan 111. I will be available at other times by appointment.

Attendance
All policies of the Tuck School apply. In addition, *unexcused* absences will lead to reduced grades as follows:
2 unexcused absences: LP
3 unexcused absences: F

Materials

Text
There is no text for this course.

The following texts may be used for reference:

Data Mining for Business Intelligence, Galit Shmueli, Nitin Patel, and Peter Bruce, Wiley, 2010.

> This is an introductory text on data mining. Although it occasionally uses advanced mathematics, most of it is accessible. It is closely integrated with the data mining software *XKMiner*.

Data Mining: Practical Machine Learning Tools and Techniques, Ian Witten, Eibe Frank, and Mark Hall, Morgan Kaufman, 2011.

>This is a very readable textbook on machine learning methods. It is written by the authors of WEKA, so it also contains a very useful guide to that open software environment.

Principles of Data Mining, David Hand, Heikki Mannila, Padhraic Smyth, MIT 2001.

>This is a more advanced text than the others cited here. Very good for theoretical understanding.

Handbook of Statistical Analysis and Data Mining Applications, Robert Nisbet, John Elder, Gary Miner, Academic Press,

>This book offers both an encyclopedic coverage of data mining and a long list of applications in the form of tutorials. It also has tutorials on a number of software packages, including Statistica, SAS Enterprise miner, and SPSS Clementine.

The following books offer insights into applications of analytics in specific domains:

Sports Data Mining, Robert Schumaker, Osama Solieman, Hsinchun Chen, Springer, 2010.

>Gives a good introduction to the various applications of analytics to sports. Many sources of data are listed. Does not go into much detail about the actual methods used.

Neural Networks in Finance, Paul McNelis, Academic Press, 2005.

>An advanced book on applications in finance.

Software
The main software used in this course is *XLMiner*. This is an Excel add-in that automates data exploration and most of the essential data mining algorithms. (*XLMiner* is owned by Frontline Systems, the makers of *Risk Solver Platform*, and will eventually be integrated into that suite.)

For data exploration and visualization we will make use of both the tools built into Excel and the tools in *XLMiner*. In addition, we will use *Spotfire*, *JMP* and *Weka* occasionally to illustrate the breadth of tools available for analytics.

Grading
Grades will be based on homework assignments, class participation, and the project. Extraordinary contributions to the intellectual process of the course will also be recognized in the final grade. The following weights will be used in grading:

Homework	*25%*
Class participation	*35%*
Project	*40%*

Schedule

Week 1
        Day 1: Introduction
        Day 2: Data exploration and visualization

Week 2
        Day 1: Data preparation
        Day 2: Performance evaluation

Week 3
        Day 1: Classification and regression trees
        Day 2: Naïve Bayes

Week 4
        Day 1: k-nearest neighbors
        Day 2: Multiple regression

Week 5
        Day 1: Time series 1
        Day 2: Time series 2

Week 6
        Day 1: Logistic regression 1
        Day 2: Logistic regression 2

Week 7
        Day 1: Neural nets 1
        Day 2: Neural nets 2

Week 8
        Day 1: Speaker 1: spatial data mining
        Day 2: Speaker 2: text mining

Week 9
        Day 1: Project presentations
        Day 2: Project presentations