# MBA8580: Business Analytics: Assignment 2

This assignment is designed to introduce you to data mining techniques for conducting association (market basket), classification (CART), and clustering analyses with XLMiner.

## Instructions

This is a team based assignment. Each group should turn in one hard copy Word document with all of its analysis. The first page should contain the names of all of your group members. You may reference an appendix with relevant XLMiner output.

See course schedule for due dates.

## Problem 1

After taking a course in Business Analytics last semester, you and your teammates decided to form a consulting firm, NOVAnalytics. Your firm was just contracted by the drugstore chain, Rite-greens. Recently, Rite-greens has experienced stagnant sales in cosmetics and is considering a variety of options including; point of sale displays, more guidance to sales personnel for promoting cross sales, and a time of purchase electronic recommendation service. The company has given you data on its past cosmetics transactions (in the file cosmetics.xls) and has asked you to conduct a market basket analysis to analyze associations among those purchases.

Develop an executive summary of your market basket analysis for upper management at Rite-greens. The summary length should be 1 page (not including the appendix).

- Be sure to justify any settings used in your analysis (e.g., support, confidence, lift ratio, etc.) and address how the results can be used by the company.
- Your summary should reference an appendix containing appropriate output from XLMiner.

## Problem 2

After NOVAnalytics' success with the Rite-greens account, an analyst at J&R Equities in Philadelphia contracted your services to help aid her study of the pharmaceutical industry. You have been asked to perform a k-means cluster analysis to better understand the structure of the pharmaceutical industry using financial data collected for 21 firms that J&R Equities is studying (provided in pharma.xls)

Prepare an executive summary of your cluster analysis findings for J&R Equities. The summary length should be 1 page (not including the appendix).

- Be sure to justify any settings used in your analysis (e.g., weights given to different variables, number of clusters formed, etc.) and address how J&R can interpret the clusters (e.g., by naming each cluster, describing patterns, etc.) to better understand the pharmaceutical industry.

- Your summary should reference an appendix containing appropriate output from XLMiner.

---

## Problem 3

You and your team at NOVAnalytics have gained particular expertise in advising small businesses that sell on eBay. After several such consulting projects you have learned that one key driver of success on eBay is having a "competitive auction", where at least two bids are placed on the item auctioned. NOVAnalytics wants to build a model that will predict whether or not any auction will be competitive so that it can advise clients on better auction settings. Your team has gathered information on 1972 eBay auctions between May-June 2004 (in the file eBayAuctions_Revised.xls) to build and test its model. Answer the following questions to guide your model development (in a 1 page document).

*Hint: Prior to building a model you will need to split the data into training (50%), validation (30%), and testing (20%) datasets.*

**a1.** Fit a classification tree with all predictors, using the best-pruned tree. To avoid overfitting set the minimum number of records in a leaf node to 50. Also, set the maximum number of levels displayed at seven (the max. allowed in XLMiner). Is this model practical for predicting the outcome of a new auction? Why or why not?

**a2.** Describe the interesting and uninteresting information that your tree's rules provide.

**b1.** Fit another classification tree (using the best-pruned tree, with a minimum number of records per leaf node =50 and seven displayed levels) this time only with predictors that can be used for predicting the outcome of a new auction. Describe the resulting tree in terms of rules.

**b2.** Based on the tree you created in b1, what can you conclude from these data about how the chances of a competitive auction relate to the auction settings set by the seller (duration, opening price, ending day, currency)? What would you recommend for a seller as the strategy that will most likely lead to a competitive auction?