

## MIS2502: Final Exam Study Guide

The exam will be a combination of multiple-choice and short-answer questions. It is a closed-book, closed-notes exam. The final exam is not cumulative.

You will not be able to use a computer during the exam. You should bring a calculator. You can use your smartphone's calculator, but that is the only thing you can use it for!

The following is a list of items that you should review in preparation for the exam. Note that not every item on this list may be on the exam, and there may be items on the exam not on this list.

### Data Mining and Data Analytics Techniques

- Explain the three data analytics techniques we covered in the course
  - Decision Trees, Clustering, and Association Rules
  - What kinds of problems can each solve? Provide a business-oriented example.
  - Make recommendations to a business based on the results of each type of analysis.
- Explain how data mining differs from OLAP analysis
  - Why would you use this instead of a data cube and a pivot table?

### Understanding Descriptive Statistics

- Be able to read and interpret a histogram
- Be able to read and interpret sample statistics

### Decision Tree Analysis

- Determine the probability of an event happening based on a series of variable values
- Why do our decision trees typically have categorical variables for the outcomes
  - For example – 0 = no default, 1 = default
- Compute and interpret the Chi-Squared statistic for a split variable.

**<<CONTINUED ON NEXT PAGE>>**

## Cluster Analysis

- Be able to read the output from a cluster analysis
  - And interpret a scatter plot of 2 dimensional data (i.e., the baseball example from the slides)
- Understand the difference between cohesion and separation
- What do you look for in the histogram that tells you a variable should not be included in the cluster analysis?
- Mean statistics – interpret root mean squared standard deviation and distance to nearest cluster
  - Relate them back to the concepts of cohesion and separation
  - What does it mean when those values are larger (or smaller)?
  - Which is better?
  - What happens to those statistics as the number of clusters increases?
    - What is the advantage of fewer clusters?
- Segment profile plot – be able to read the histograms for each segment
  - For a particular segment, be able to tell if the values tend to be higher or lower than the average for the entire data set

## Association Rules (Market Basket Analysis)

- Be able to read and interpret the output from an association rule analysis
  - Find the strongest (or weakest) rule from a set of output
- Understand the difference between confidence, lift, and support
  - You should be able to explain the difference between them
    - Can you have high confidence and low lift?
  - You won't have to compute them, but you should understand how they are computed so you can interpret the statistics
- What's the difference between a standard association rule analysis and a sequence analysis?
  - What situations would a sequence analysis make sense? When wouldn't it?
- Given a set of baskets, compute and interpret confidence, support, and lift for an association rule.
- Given a table of aggregate purchase numbers for two products, compute and interpret the lift for the rule that includes those two products.

## Data Visualization

- Be able to assess an infographic or chart by applying data visualization principles.
  - Tell a story
  - Graphical integrity (lie factor)
  - Minimize graphical complexity (data ink)
- Explain how a visualization can be improved based on those principles.