

MSIS 6300 – BA for Managers

Data Minign Assignments

In this assignment you will choose to use a free/open-source data mining tool, KNIME (<http://knime.org>). You are to analyze a given dataset (about the voting behavior of a number of counties in the U.S.) to develop and compare at least three different types of prediction (i.e., classification) methods that predicts whether a county will say "yes" or "no" to legalizing gaming at the ballot). Here are the specifics for this assignment:

- Use the following tools
 - KNIME (download and install on a PC/Laptop).
- Download “Voting Behavior” data and the brief data description from the D2L
 - The data is given in MS Excel format.
- Follow the 6 steps in CRISP-DM process model
 - Understand the domain and the problem you are trying to solve (via literature).
 - Understand, and preprocess the data (be very critical about the data).
 - Develop at least three classification models (e.g., Decision Tree, SVM, ANN, etc.).
 - Compare the accuracy results (use confusion matrixes and ROC curves).
- Present your results in an organized report
 - Include a cover page.
 - Write an “Executive Summary” (1 page long).
 - Use the 6 steps in CRISP-DM to organize the remainder of the report.
 - Include a conclusion page, where you need to comment on the tool and techniques you’ve used. What was good and what was bad, etc.
 - Make sure to integrate figures (graphs, charts, tables, screen-shots) into the text as you see necessary. Do not use Appendixes.
 - Try not to exceed 15 pages in total, including the cover (use 12 point Times New Roman fonts, and 1.5 line spacing).

Deadline:

- Submit/upload your report by the due date/time stated on the specific D2L/Dropbox folder.
- The report should be uploaded as a single Microsoft Word Document.