Instructions for the Research Project

The research project will materialize everything we will study during the course. You will feel the application of the concepts we are studying.

I. Data Analysis Project.

1. Identify the problem(s) to be solved or opportunities to be realized by mining the selected data set.
2. Consider the following data preparation questions and explain your answers. When appropriate cite resources that support your answer. Explain how the answers, and data preparation, differed when you chose a different data mining method.
    a. Should instances with missing values be deleted?
    b. Should missing values be specially coded and then retained in the data set?
    c. Should numeric values be assigned predetermined ranges or left for the algorithm to split?
    d. Should categorical variables be grouped or coded to reflect a hierarchy?
3. To explore the problem or opportunity, use two or more of the following data mining methods covered by this course:
    a. regression: linear regression, discriminant analysis or logistic regression,
    b. decision trees,
    c. neural networks,
    d. hierarchical or k-means clustering,
    e. association rules,
    f. time series,
    g. genetic algorithms.
4. Describe the algorithms chosen, and indicate why you chose them. Exploring a method of interest is a satisfactory reason for this course paper.
5. Explain how and why you used specific pruning parameters or other adjustments to create a sparser model.
6. Compare the alternative solutions using methods found in comparative studies in the literature. For example, see "**Data mining for network intrusion detection: A comparison of alternative methods**" *Dan Zhu*, *G Premkumar*, *Xiaoning Zhang*, *Chao-Hsien Chu*. *Decision Sciences*.Atlanta: Fall 2001.Vol.32. http://www.findarticles.com/p/articles/mi_qa3713/is_200110/ai_n8954240 Report the results of the accuracy measures available with the software. If the software used does not have built-in accuracy reporting then manually test the model's accuracy on a small hold-out test sample of the data. The hold-out method creates separate training and test sets. This is particularly useful when testing the model on data from a later time period.
7. Create a table showing the number of cases correctly identified, Type I, and Type II errors. In addition, a ROC curve is appropriate with discriminant analysis and logistic regression. For these methods, changing the parameters for the line separating the classes, changes the percentages of Type I and Type II errors. Medical practitioners like ROC curves because they show the tradeoff between false positives and false negatives.
8. Which data mining method(s) seem superior for the chosen data set? Did the method that performed best in your study also dominate in similar comparative studies?
9. Compare the results or recommendations that would result from the use of the different methods.
10. Based on your analysis, justify a conclusion or recommendation.
11. Cite the relevant literature using APA formatting described at http://www.umuc.edu/library/libhow/citeright_tutorial_apa_articles.cfm Any publication listed in the references should be cited in the paper.
12. Organize the paper into the sections of a formal research paper: Introduction, Methods, Results, etc., Use the resources provided in the Effective Writting Center: http://www.umuc.edu/writingcenter/

II. Writing Skills Research Paper

1. Writing skills are critically important to succeed in the Graduate School (TGS) and in your future careers. If conducting research is new to you, consider using the library's helpful resources. To learn about effective Internet research, consult the Information and Library Services (ILS) http://www.umuc.edu/writingcenter/onlineguide/index.cfm
2. Before you search the Web, read

   http://www.umuc.edu/writingcenter/writingresources/sources.cfm for Web sites provided by ILS.

3. Effective Writing Center(EWC) provides number of helpful tool to improve your writings, including Plagiarism Tutorial. It's paramount you check this tutorial before start writing.
4. Before locating a full-text journal article in a database, read the http://www.umuc.edu/library/libhow/ . Then go to the Computer/Information Science topic area in the Resources by Topic section of the Library Databases and E-Journals page to select a database in which to search. The database ACM Digital Library is a good one to start with.
5. Use the APA style guide for your citations. After each title add a short note evaluating its quality as a research tool (back to evaluation criteria), and its quality relative to the other sources cited in your bibliography.

III. Original Work and Plagiarism.

Policy on Academic Integrity: Please review the UMUC policy on academic dishonesty and plagiarism. Your paper has to be in complete compliance with UMUC's zero-tolerance policy regarding plagiarism. "Plagiarism includes, but is not limited to the following: copying verbatim all or part of another's written work; using phrases, charts, figures, illustrations, or mathematical or scientific solutions without citing the source; paraphrasing ideas, conclusions, or research without citing the source; and using all or part of a literary plot, poem, film, musical score, or other artistic product without attributing the work to its creator. Students can avoid unintentional plagiarism by carefully accepted scholarly practices. Notes taken for papers and research projects should accurately record sources of material to be cited, quoted, paraphrased, or summarized, and papers should acknowledge these sources in footnotes." Have in mind that all papers will be submitted to *Turnitin* prior to being read and graded.
Also note that any graded assignments must be the student's own work and original for this course. Work prepared for other courses or use of material obtained for this course from other students, past or present, is expressly prohibited and can result in a grade of zero for an assignment.
Please review the syllabus and pay attention on the requirements for original work and plagiarism.

IV. Next step.

Please select your topic for research project and post 1-2 paragraphs summary (abstract) on your intended topic as **a New topic** in this Conference. Please change the title of your post with **the title** of your project.