CS 635 Project Guidelines

A **Research Paper** of significant value to the field of Biomedical Informatics will be written. Topics of interest to pairs of students from different backgrounds may be proposed by the students and approved by the instructor. It should be based on data from current publication like micro-array data from the *Gene Expression Omnibus (GEO)* or a medical data set likes NHANES. Choose a publication so that at least one student is familiar with the biological problem and one student is familiar with the statistical/machine learning algorithm. Recommended algorithms for CS 635 are PCA, Neural Nets, Decision Trees, SVM, etc. For instance, if you are interested in liver cancer and Neural Nets, search GEO for both and select a publication from there. You may chose any professional software that is appropriate, you don't need to use the software used in the publication – see list on Blackboard. However, Weka is considered teaching software and not an option.

Make sure that you understand the biological problem in the selected publication and the algorithms completely. You might need additional resources besides the textbooks. Some additional books are listed on Blackboard (external links). Your paper can only be as good as your understanding of the material. Start early with the project, there might be unexpected problems in the process. See the instructor early in the semester; by end of the semester office hours are very crowded.

The paper should follow the JAMIA guidelines (http://jamia.oxfordjournals.org/for_authors/index.html) (Case Study, 5-10 pages, 1,500-3,000 words) or Bioinformatics guidelines (http://www.oxfordjournals.org/bioinformatics/for_authors/submission_online.html).

For Bioinformatics see also http://www.oxfordjournals.org/bioinformatics/for_authors/word-template.zip .

Source code (e.g. R-scripts) should be submitted separately and does not count for the word limit.

Here is a possible outline of the paper:

1. Introduction - A short statement that presents the purpose of the paper.
2. Background - Motivate and place the work in context.
   Subsections might address the significance of the problem being addressed, other approaches to the problem, and prior work on the approach being reported.
3. Formulation process - the biological experiment design.
4. Model description – describe the statistical or machine learning model
5. Analysis – describe the authors statistical/ML analysis, and what you did different and why.
6. Results – Report your results and potential differences to the authors' results.
7. Discussion - Subsections might address significance, limitations, and problems with the analysis.

8. Conclusion - Short summary of the model and implications.
9. References – the paper and any additional material you used.

Please see or email the instructor immediately, if any problems with the analysis or the data arise.

STANDARDS OF EVALUATION

A = insightful

- Description of the publication's goals and importance demonstrates reflection and learning beyond just summarizing what was written in the publication.
- Application to the analysis demonstrates real ownership of the ideas.
- Tight logical link between the problem description and analysis in the paper, the concepts and algorithms from the course, and the conclusions or implications drawn.
- Conclusions are a logical outgrowth of applying the concepts to the description of the paper.
- The conclusions/implications regarding the algorithms go beyond the boundaries of our classroom discussion, and show some initiative in exploring the learning opportunity.
- Focus of the paper is depth of understanding of the biological problem and the algorithms used in the publication, rather than breadth of coverage of some unimportant detail.
- Results are illustrated in an appropriate visual/graphical form.
- Paper probably could be used as the basis for a nice talk on the featured algorithm or problem.

B = competent

- Paper demonstrates correct and complete (but not insightful) analysis and application of the course concepts to the referred publication.
- It is clear from reading the paper that the student understands what was communicated and discussed in the publication, but probably has not necessarily learned something more from the analysis/application and reflection/write-up of what the real problems and implications of the publication are. (*NOTE: this may result in very little feedback from the grader*.)
- Paper's link between problem description, concepts, and applications is not tight.
- Not clear how the concepts relate to the description, or how the description and concepts logically lead to the conclusions or implications.
- Paper touches on several concepts or algorithms on the surface, rather than exploring any one of them in any useful depth or trying different parameter settings to match the results of the paper.
- Paper missed some of the subtleties of the featured publication.

C (or worse) = Inadequate

- Description of the publication's key issues is incomplete or incorrect.
- Application of the algorithm(s) in the publication is flawed.
- Paper does not link the publication's problem description, algorithms, and application.
- Paper describes the publication without really using the algorithms to analyze the problem.
- Paper bounces around among a variety of algorithms without fully explaining or applying any of them.

- Paper uses words/terms from the publication without communicating a sense of understanding of their real meaning.
- Paper doesn't make clear if there were issues with the publication, if and how the student resolved them, or that the student dealt with them at all.