CSE 598E/ STAT 598A  
Project Three  
Due 30 April 2007

For this project, you will work as individuals. You are to write a 5-10 page summary of a relevant research paper in the literature on Data Mining or Machine Learning and give a short (10 minute) presentation on it during the last week of class (April 30-May 4). The paper should be published in 1999 or later. Possible sources include the bibliography of the text book and the web pages of the following faculty, Jia Li (www.stat.psu.edu/~jiali), Lee Giles (http://clgiles.ist.psu.edu/), and Hongyuan Zha (now at GA Tech)(http://www.cse.psu.edu/~zha/papers.html). Also recommended are the web pages of text book authors T. Hastie, R. Tibshirani, and J. Friedman (you can google them, they all work at Stanford) and two people who gave colloquia here this semester, Daniel Kifer (http://www.cs.cornell.edu/people/dkifer/) and Jinmeng Sun (http://www.cs.cmu.edu/~jimeng/). There are many, many other sources—e.g., the SIAM conference on Data Mining, IEEE International Conference on Data Mining, and the International Conference on Machine Learning, (ICML).

Papers about data mining or machine learning applications in image processing, compilers, web search engines, program debugging or your favorite area of science or engineering are most welcome as long as there is sufficient data mining or machine learning content. The paper you choose cannot be a paper on which you or any other class member or the instructor is a co-author.