Fall 2009 Course Announcement

6K:275 Knowledge Discovery
Prof. Nick Street, Management Sciences Dept.
2:30 - 3:45 TTh
104 EPB

Gathering large volumes of data is both easy and cheap. Turning that data into new, useful knowledge is much more challenging. In this course you will be introduced to the emerging field of Knowledge Discovery in Databases (KDD), also called Data Mining. KDD combines ideas from machine learning / artificial intelligence, database analysis, and statistics to locate and verify patterns in large data stores. From diagnosing disease to detecting credit card fraud to making NBA coaching decisions – any time data are plentiful and expertise is rare and expensive, KDD can help.

The course has no specific prerequisites but some knowledge of probability and statistics (conditional probability, hypothesis testing) and basic calculus will be assumed. The course is appropriate for graduate students from any discipline and is specifically designed to be accessible to non-technical students while still being conceptually challenging for IS/CS/Engineering students. No computer programming is required. All the assignments will have you digging new knowledge from a dataset of your own choosing, so if you have a pet problem please bring it along.

Specific topics will include:

- The knowledge discovery process: data preparation, cleaning, and transformation; model evaluation; knowledge integration

- Predictive modeling
  - statistical methods: nearest neighbor, Bayesian learning, linear models
  - decision trees
  - artificial neural networks
  - support vector machines
  - ensemble methods

- Learning issues
  - performance evaluation: accuracy, sensitivity, etc.
  - overfitting avoidance
  - cross-validation and bootstrapping
  - bias vs. variance
  - dimensionality reduction: feature selection, principal component analysis, etc.

- Descriptive modeling
  - clustering: k-means, mixture models, agglomerative / divisive clustering
  - association rule learning
  - outlier detection

- Web mining and recommender systems