Goals:
The objective of knowledge discovery and data mining process is to extract nontrivial, implicit, previously unknown, and potentially useful information from massive datasets. The course is intended to serve as an introduction to the fundamental techniques required to support this process. The course is structured to provide ample opportunity for participants to learn about a growing new research area, and scout around for promising research topics by a hands-on experience.

Prerequisites:
Basic knowledge in Data Base Systems (CIS 616); programming skills in C or C++; basic statistics, graph theory, and linear algebra.

Texts:

Topics: will be tailored to interests of the participants. Content will include:
I. An overview of data mining tasks and techniques;
II. Data preprocessing:
   (1) data cleaning and transformation,
   (2) data reduction.
III. Core data modeling topics:
   (1) model functions (classification, regression, clustering, summarization, sequence analysis, outliers analysis),
   (2) model representation (decision trees, Bayesian belief networks, neural networks, density models, partitioning, hierarchical, density-based, grid-based and model-based clustering algorithms, Apriori algorithm, correlation analysis).
IV. Advanced topics
   (1) spatial data mining,
   (2) temporal and sequence data mining,
   (3) text and web mining,
V. Reading and research projects presentations.

Grading: Homework, reading assignments and an individual research project.