# Syllabus

## CSGA 6950: Applied Algorithms for Data Analysis

### Department of Computer and Information Sciences
### Fordham University
### Fall 2005

### Course Description:

This course will cover data mining and machine learning algorithms for analyzing large data sets as well as the practical issues that arise when applying these algorithms to real-world problems. It will balance theory and practice—the principles of data mining methods will be discussed but students will also acquire hands-on experience using state-of-the-art data mining software (SAS Enterprise Miner, C5.0) to solve scientific and business problems. Students will learn about data mining algorithms for: classification and prediction (decision trees, neural networks, nearest-neighbor, genetic algorithms, Naïve Bayes), clustering (K-means), association rule mining (Apriori) and algorithms for handling complex data types (text-mining, image-mining, etc.). In addition, the process for mining/analyzing data will be covered, including the following issues: data warehousing and OLAP, data preprocessing, data transformation, and model evaluation. Textbook readings will be supplemented with current articles on data mining technology and applications. Each student will also, with the aid of the instructor, select and complete an application-oriented or research-oriented course project.

### Instructor: Dr. Gary Weiss


- **Office:** JMH 328A
- **Phone:** 718-817-0785
- **Email:** gweiss@cis.fordham.edu
- **Office Hours:** Tuesday, Friday 11-12; 1:30-2:15


### Course Objective:

To develop an understanding and familiarity with data analysis and data mining algorithms and be able to apply them to solve real-world problems.

### Textbook:

*Introduction to Data Mining:* by Tan, Steinbach and Kumar (2005)

There will also be supplemental readings assigned during the semester.

### Grading:

There will be one exam, a midterm, during the course (there will be no final exam since the main focus after the midterm will be on the course projects). Homework assignments are an important part of the class and should be completed on time. Readings are also expected to be completed on time and class participation is an important component of this class. A course project, selected by each student (or team of two students) and the instructor will be a key component of your final grade. The percentages given below are guidelines and minor modifications may be made as needed to reflect circumstances in the course. Each student may also be asked to present a technical paper.

| | |
|---|---|
| Homework | 25% |
| Participation & Presentations | 10% |
| Midterm | 25% |
| Final Project | 40% |

**Academic Honesty:**

All work produced in this course should be your own unless specifically allowed. Violations of this policy will be handled in accordance with university policy which can include automatic failure of the assignment and/or failure of the course. In situations where collaboration is permitted or required you should be careful to cite any individual who provided assistance and is not already credited on the work.

**Topics to be Covered**

1. Introduction to Data Mining
2. Data
3. Data Exploration
4. Classification I: Basic Concepts, Decision Trees and Model Evaluation
5. Classification II: Advanced Techniques (Rule-Based learners, Neural Networks, Bayesian Learning, etc.).
6. Classification III: Utility-Based Data Mining
7. Association Analysis
8. Cluster Analysis
9. Data Mining Applications
10. Advanced Topics: Text Mining/Retrieval, Link Analysis, Image Mining, etc.

In addition to the topics listed above, there will be two technical talks on related topics from outside researchers. At least one class will be devoted to student presentations of their course projects.