

Course Manager – Dr. Nazli Goharian, Clinical Assistant Professor

3 credit hours; elective for CS & CPE; 150 min. lecture each week

● **Current Catalog Description**

This course will provide an introductory look at concepts and techniques in the field of data mining. After covering the introduction and terminologies to Data Mining, the techniques used to explore the large quantities of data for the discovery of meaningful rules and knowledge such as market basket analysis, nearest neighbor, decision trees, neural networks, and clustering are covered. The students learn the material by implementing different techniques throughout the semester (3-0-3).

**Textbook**

J. Han, M. Kamber. Data Mining Concepts and Techniques, Morgan Kaufmann

**Course Goals** - Students should be able to:

- Explain the Data Mining motivation and applications.
- Explain the Data Mining Architecture.
- Explain Data Preprocessing motivation and techniques.
- Explain various Data Mining algorithms such as Naïve Bayes, Neural Networks, Decision Tree, Association-Rules, and Clustering.
- Explain the scalability issues for each of the algorithms discussed in the class and how they can be modified for scalability.
- Design and implement data mining systems using various data pre-processing techniques and mining algorithms.
- Apply the research ideas into their experiments in building data mining systems.

**Prerequisites by Topic**

- Data Structures, Algorithm and Strong Object Oriented Programming.

**Major Topics Covered in Course**

1. Introduction: (basic terminology, motivations, applications, Knowledge and Data Discovery process (KDD), Data Warehouse, Data Mining architecture, Supervised vs unsupervised algorithms.	
2. Data Preprocessing	
3. Naïve Bayes algorithm	
4. Neural Network algorithm	
5. Decision Tree algorithm, including some scalable algorithms such as SLIQ,....	
6. Association Rules	
7. Clustering	
9. Implementations and scalability issues	
11. Paper presentation	
Tests (1,2,3)	

**Laboratory projects (specify number of weeks on each)**

N/A

**Estimate CSAB Category Content in Credit Hours**

	CORE	ADVANCED			CORE
Data Structures		1		Computer Organization and Architecture	
Algorithms		1		Concepts of Programming Languages	
Software Design		1			

**Oral and Written Communications** - Every student is required to submit at least 5 written reports (not including exams, tests, quizzes, or commented programs) of typically 3 pages and to make 1 oral presentations of typically 20 minutes duration. Include only material that is graded for grammar, spelling, style, and so forth, as well as for technical content, completeness, and accuracy.

- Design and Summarization of the prototype implementation.
- Research paper presentation

**Social and Ethical Issues** - Please list the topics that address the social and ethical implications of computing covered in all course sections. Estimate the class time spent on each topic. In what ways are the students in this course graded on their understanding of these topics (e.g., test questions, essays, oral presentations, and so forth)?

- Ethical issues in mining the data. 1 hour.

**Theoretical Foundations** - Please list the types of theoretical material covered, and estimate the time devoted to such coverage in contact (lecture and lab) hours.

- Mining algorithms 15
- scalability: 4 hrs.
- Advanced and research related material. 6

**Problem Analysis** - Please describe the problem analysis experiences common to all course sections.

- Analysis of the experimental results obtained by the implementation of different algorithms and techniques in the data mining class projects.

**Solution Design** - Please describe the design experiences common to all course sections.

- Design modules for implementation of various data preprocessing and algorithms, and scalability techniques for a data mining system.

**Other Course Information**

- **Additional Suggested Course Assignments**

None.

- **Planned Course Enhancements:**

None.