

İSTANBUL KÜLTÜR UNIVERSITY
DEPARTMENT OF INDUSTRIAL ENGINEERING
IE 020 - INTRODUCTION TO DATA MINING
SPRING 2010

Instructor: Assist. Prof. Fadime ÜNEY-YÜKSEKTEPE **E-mail:** f.yuksektepe@iku.edu.tr

Office: A205 **Office hours:** Thursday 14:00 – 15:30 or by appointment

Class hours: Thursday 09:00 – 11:45 **Room:** D-257

URL address for the course material: <http://sites.google.com/site/fadimeyuksektepe/ie-020>
<http://ie.iku.edu.tr/coursehome.asp?CourseID=826&PeriodID=44>

Textbooks:

- P.N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, Pearson Education, 2005.
- C. Shmueli, N. R. Patel, P. C. Bruce, *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, Wiley-Interscience, 2007.
- R. J. Roiger and M. W. Geatz, *Data Mining: A Tutorial Based Primer*, Addison Wesley, Pearson Education, 2003.

Recommended Reading:

- G. Loveman, *Diamonds in the Data Mine*, Harvard Business Review, pp. 109-113, May 2003.
- R. Gerritsen, *Assessing Loan Risks: A Data Mining Case Study*, IT Pro, pp.16-21, November-December 1999.
- M. Weatherford, *Mining Fraud*, IEEE Intelligent Systems, pp. 4-7, July-August 2002.

Course Description: Due to recent advances in sophisticated hardware and software technologies, large quantities of data can be acquired, processed and stored. However, the amount of collected data frequently increases and constitutes large complicated databases. As a result of these structures, data mining nowadays receive considerable attention. Data Mining is the process of investigating and extracting implicit, previously unknown and potentially useful information from large data by using one or more computer-based learning techniques. Many different data mining methods exist; for example classification, association analysis, clustering, anomaly detection, feature selection. At the end of this course, students are expected to possess the fundamental skills needed to conduct research in data mining or to apply data mining techniques to real-world applications.

Evaluation

Midterm Exam	25%	Homework	10%
Project	25%	Final exam	40%

Project Guidelines:

A term project that covers a potential or real business problem will provide opportunity of facing the challenge of making business use of data mining. A number of alternatives will be provided by the instructor. A project proposal that describes the setting and the goals of the project should be submitted by November 25, 2010. The final report, which details the project setting, the models built (at least four data mining tools should be compared in terms of functionality for the chosen application), relevant testing and evaluation, and conclusion is due December 30, 2010. A short presentation of the project will be made in class. Course projects should be done in groups of 2.

COURSE OUTLINE

Week	Date	Topic
1	September 30	Chapter 1 – DATA MINING Data Mining: Definitions, Motivation, Concepts and Techniques
2	October 7	Chapter 2 – DATA PREPROCESSING Data Preprocessing: Data Sampling and Data Cleaning
3	October 14	Chapter 2 – DATA PREPROCESSING cont.'d Data Preprocessing: Feature Selection and Dimensionality Reduction
4	October 21	Chapter 4 & 5 - CLASSIFICATION Tree-based, Rule-based and Instance-based Methods, Bayesian Methods
5	October 28	Chapter 4 & 5 - CLASSIFICATION cont.'d Neural Networks, Linear Discriminant Analysis, Support Vector Machines
6	November 4	Chapter 4 & 5 - CLASSIFICATION cont.'d Ensemble Methods and Model Evaluation
7	November 11	DATA MINING SOFTWARE: WEKA
8	November 18	NO CLASSES (KURBAN BAYRAMI)
9	November 25	MIDTERM
10	December 2	Chapter 6 & 7 – ASSOCIATION ANALYSIS Apriori Algorithm and its extensions, Pattern Evaluation
11	December 9	Chapter 6 & 7 – ASSOCIATION ANALYSIS Sequential Patterns and Graph Mining
12	December 16	Chapter 8 & 9 - CLUSTERING Partial and Hierarchical Clustering Methods, Graph-based and Density-based Methods, Cluster Evaluation
13	December 23	Chapter 10 – ANOMALY DETECTION Approaches to Anomaly Detection
14	December 30	PROJECT PRESENTATIONS