

BIL 713 - Data Mining

Fall 2014

Course Description

Knowledge Discovery and Data Mining (KDD) is an interdisciplinary area which utilises various techniques from statistics and machine learning to extract useful information from large amounts of data. In this course, we will discuss the main concepts and techniques related to processing and evaluating data. Throughout the semester, the students will be exposed to various problems in data mining and the corresponding potential solution proposals.

Detailed Information

Instructor: Dr. Nazlı İközler-Cinbiş • nazli@cs.hacettepe.edu.tr • tel: 297 7500 – 147

Schedule: Wednesdays at 13:00-16:00.

Office Hours: By appointment.

Textbook:

Introduction to Data Mining, P.-N. Tan, M. Steinbach, and V. Kumar, Addison Wesley, 2005.

Supplementary Books:

- Data Mining: Concepts and Techniques, 2nd ed., Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2006
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, T. Hastie, R. Tibshirani, and J. Friedman, Springer-Verlag, 2001.

Webpage: <http://web.cs.hacettepe.edu.tr/nazli/courses/bil713/index.html>

Communication: The course webpage will be updated regularly throughout the semester with lecture notes, assignments and important updates. All other communications will be carried out through the Piazza system:

<https://piazza.com/hacettepe.edu.tr/fall2014/bil713>.

Topic List:

- 1 Introduction
 - 2 Data Preprocessing
 - 3 Finding Similar Entities
 - 4 Classification and Model Evaluation
 - 5 Clustering
 - 6 Frequent Itemset Mining
 - 7 Association Rule Mining
 - 8 Sequence Mining
 - 9 Anomaly Detection
 - 10 Text and Web Mining
-

Grading (Tentative):

- Survey paper + Presentation (30%)
- Project (40%)
- Final exam (30%).

Survey Paper: The students are expected to select a specific topic on data mining and write a short survey paper which reviews the recent literature, i.e. the major works over the last years, on that topic. The students are expected to summarize the selected topic and talk about two recent algorithms on that topic in a short presentation in class.

Project: In the context of the course, the students are expected to carry out an in-depth project on data mining. This project will involve the detailed analysis and implementation of an approach to an existing problem of data mining.

Important Dates:

Survey Topic Selection	8th October 2014 (<i>tentative</i>)
Project Topic Selection	15th October 2014 (<i>tentative</i>)
Project Submission	17 December 2014 (<i>tentative</i>)

Academic Integrity:

The students are expected to strictly adhere to the academic integrity policy of the Department, act honestly and respect the rights of the others in carrying out all academic assignments. Academic dishonesty, including cheating, fabrication and plagiarism will not be tolerated. Unless stated otherwise, all the work on the assignments must be carried out individually. While the discussions over the general concepts about the course are allowed, discussions related to a particular solution to a specific problem, (actual or pseudo) code sharing are strictly forbidden. Using assistance from the internet without providing proper citation is also considered as violation of the academic integrity. Note that, all the aforementioned violations to academic integrity policy are subject to disciplinary action.