IE 425

DATA MINING

SPRING 2021

Instructor: Necati Aras, Ph.D.
Office: Engineering Building 4070
Phone: 358-7506
E-mail: arasn@boun.edu.tr
Web Site: www.ie.boun.edu.tr/~aras

Data that are needed for making managerial decisions is accumulating at an increasing rate due to a number of technological advances. As a result of innovations such as the internet, electronic banking, point-of-sale devices, barcode readers and e-tailers, electronic data collection has turned out to be inexpensive. Consequently, data warehouses and data marts designed for managerial decision support contain huge amounts of data. Data mining that evolved from the disciplines of statistics and artificial intelligence is concerned with applying various techniques to make intelligent use of data repositories. There have been several successful applications in areas such as credit rating, database marketing, fraud detection, stock market investments and customer relationship management.

This course will examine methods that proved to be useful in recognizing patterns and making predictions. We will review applications and provide an opportunity for hand-on experimentation with data mining algorithms. At the end of the course students will have developed an understanding of the strengths and limitations of popular methods to solve regression and classification problems in data mining.

Textbook

An Introduction to Statistical Learning with Applications in R written by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, ISBN 978-1-4614-7137-0 ISBN 978-1-4614-7138-7 (eBook)

Tentative Course Schedule:

Week	TOPIC
1	Introduction to Data Mining
2	Supervised Methods: Classification Problems Classification Trees
3	Classification Trees, Performance Evaluation
4	Regression Trees, Performance Evaluation
5	Ensemble Methods: Bagging, Random Forests
6	Naïve Bayes Approach, K-nearest Neighbor Classification
7	Simple and Multiple Linear Regression
8	Logistic Regression
9	Midterm Exam
10	Spring Break
11	Unsupervised Methods: Clustering, Hierarchical Clustering
12	K-means Clustering
13	Association Analysis

Evaluation:
Assignments (3): 30%
Midterm Exam: 25%
Final Exam: 25%
Project: 20%: