

# FE 581

## DATA MINING

SPRING 2021

**Instructor: Necati Aras, Ph.D.**

**Office:** Engineering Building 4070

**Phone:** 358-7506

**E-mail:** [arasn@boun.edu.tr](mailto:arasn@boun.edu.tr)

**Web Site:** [www.ie.boun.edu.tr/~aras](http://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 hands-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 data mining techniques.

### **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:**

	<b>TOPIC</b>
Week 1 (Mar. 22)	Introduction to Data Mining
Week 2 (Mar. 29)	Supervised Methods: Classification and Regression Problems
Week 3 (Apr. 5)	Classification Trees, Regression Trees
Week 4 (Apr. 12)	Classification Trees, Regression Trees
Week 5 (Apr. 19)	Ensemble Methods: Bagging, Random Forests
Week 6 (Apr. 26)	Ensemble Methods: Bagging, Random Forests
Week 7 (May 3)	Naïve Bayes Approach, K-nearest Neighbor Classification
Week 8 (May 10)	Spring Break
Week 8 (May 17)	Multiple Linear Regression
Week 9 (May 24)	Midterm Exam
Week 10 (May 31)	Logistic Regression
Week 11 (Jun. 7)	Unsupervised Methods: Hierarcical Clustering
Week 12 (Jun. 14)	K-means Clustering
Week 13 (May 6)	Assocation Analysis
Week 14 (May 13)	Principal Component Analysis

**Evaluation:**

Assignments (2): 20%

Midterm Exam: 30%

Final Exam: 30%

Project: 20%