# ETM 580

# DATA MINING

## FALL 2020

**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 data mining techniques.

**Textbooks**

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 (October 28) | Introduction to Data Mining |
| 2 (November 4) | Supervised Learning: Classification and Regression Problems |
| 3 (November 11) | Classification Trees, Regression Trees |
| 4 (November 18) | Classification Trees, Regression Trees |
| 5 (November 25) | Ensemble Methods: Bagging, Random Forests |
| 6 (December 2) | Naïve Bayes Approach, K-nearest Neighbor Classification |
| 7 (December 9) | Multiple Linear Regression |
| 8 (December 16) | Logistic Regression |
| 9 (December 23) | Logistic Regression |
| 10 (December 30) | Unsupervised Learning Method: Clustering |
| 11 (January 6) | Hierarchical Clustering |
| 12 (December 13) | K-means Clustering |
| 13 (December 20) | Association Rule Mining |

**Evaluation:**
Assignments: 25%
Midterm Exam: 25%
Final Exam : 25%
Project: 25%