

## **IE 451 – Data Visualization and Analysis Course Syllabus**

**Course Instructor:** Kemal Kılıç  
**Office:** FENS 1032  
**Tel:** 216- 483 9596  
**E-mail:** [kkilic@sabanciuniv.edu](mailto:kkilic@sabanciuniv.edu)

### **Course Description:**

The course will address unsupervised learning, supervised learning, association rule mining and feature subset selection problems and introduce various techniques proposed as solutions and present their implementation particularly in the context of operations management. Data Visualization will also be introduced as part of the curriculum.

Among others, probabilistic and statistical methods, clustering algorithms, classification algorithms, multiple linear regression, a priori algorithm, metaheuristics (such as genetic algorithms, simulated annealing, etc.) in the context of feature subset selection will be covered as part of the toolbox that are widely utilized in data mining.

### **Course Web Site: *SuCourse***

Lecture materials, including the class overheads, readings, assignments etc. will be available at the course web site prior to the lectures. Students are expected to check the web site regularly in order to attain the recently posted material.

### **Reference Texts:**

TBA. Please check the *SuCourse*.

### **Marking Scheme**

Assignments	20%
Midterm I	25%
Midterm II	25%
Final	30%

### **Assignments**

There will be four *group* assignments. Questions will be mostly implementation of the algorithms with a software (e.g. WEKA).

### **Objection Policy**

Concerns regarding marks will be accepted until two days before the last date of grades submission

### **Academic Conduct**

**Do not plagiarize other people's work. Students should be aware that anyone who engages in actions prohibited by the University's policy on academic honesty will be subject to disciplinary action.**

### **Course Content**

1. Introduction
2. Validation
  - a. Metrics for classifier evaluation

- b. Methods for classifier evaluation
- c. Costs in data mining
- d. Bias – Variance Tradeoff
- 3. Classification
  - a. 1 – R
  - b. Naive Bayes
  - c. Decision Trees
  - d. Instance Based Learning
  - e. Logistic Regression
- 4. Regression
  - a. Multiple Linear Regression
- 5. Clustering
  - a. Partitional Clustering: K-Means / K-Medians / K-Medoids
  - b. Hierarchical Clustering
  - c. Probabilistic and Generative Clustering: Expectation – Maximization Algorithm
  - d. Density Based Clustering: DBSCAN
- 6. Association Rule Mining
  - a. Basic Terminology
  - b. *A priori* Algorithm
- 7. Data Understanding
  - a. Descriptive Analytics
  - b. Data Visualization
- 8. Data Preparation
  - a. Tidy Data
  - b. Outliers
  - c. Imbalanced Data
  - d. Missing Values
  - e. Normalization
- 9. Meta Heuristics
  - a. Local Search
  - b. Simulated Annealing
  - c. Genetic Algorithms
  - d. Tabu Search
  - e. Beam Search
- 10. Feature Subset Selection
  - a. Filtering Methods
  - b. Wrappers
  - c. Embedded Techniques
  - d. Feature Reduction, e.g. Principal Component Analysis
- 11. Neural Networks

### **Course Schedule and Disclaimer**

The course schedule can be found in ScheduleFall2021\_IE451.xls which will be posted to the SuCourse. However, the instructor reserves the right, when necessary, to change examination

dates, and modify the syllabus and course content. Modifications will be announced in class. Students are responsible for announced changes.