

IE 525 - Operations Research and Data Mining Course Syllabus (Spring 2024)

Course Instructor: Kemal Kılıç
Office: FENS 1032
Tel: 216- 483 9596
E-mail: kkilic@sabanciuniv.edu

Course Description:

The course will address unsupervised learning, supervised learning, association rule mining and feature subset selection problems, focus on the optimization formulations of these problems, discuss various techniques proposed as solutions and present their implementation particularly in the context of operations management.

Among others, probabilistic and statistical methods, possibilistic methods, clustering algorithms, decision trees, metaheuristics (such as genetic algorithms, simulated annealing, etc.) and mathematical programming will be covered as part of the toolbox that are widely utilized in data mining.

The course will include case studies from both manufacturing and service industries.

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

Two Exams	65% (25% + 40%)
Term Project	
- Proposal	5%
- Progress Report	10%
- Final Presentation & Report	20% (5%+15%)

Assignments

There will be four assignments. Questions will be regarding to the theoretical foundations and might also require applications with a software.

Term Project

Students should submit a term project which will be essential part of their evaluation. Possible term projects can be (but not limited to) an extensive review of literature for a particular problem, a solution proposal for a particular problem (which includes a literature review to a degree, but on top of it provide solution proposals and slight validation of the proposed methodology) or practical application of a data mining software to a particular problem (which again includes an extensive literature review as well as discussion on the result of the analysis).

Students should submit three reports (proposal, progress report and final report) throughout the semester and deliver a 20 minutes presentation at the end of the semester.

Final Report is due to TBA.

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
3. Classification
 - a. 0-R/1 –R
 - b. Naive Bayes
 - c. Decision Trees
 - d. Instance Based Learning
 - e. Logistic Regression
 - f. Support Vector Machines
 - g. Ensemble Learning (Bagging, Randomization (e.g., Random Forest), Boosting (e.g. AdaBoost), Stacking)
4. Regression
 - a. Multiple Linear Regression
 - b. Regularization in MLR (e.g., Lasso & Ridge)
 - c. Classification and Regression Trees (CART) & KNN Regression
5. Clustering
 - a. Partitional Clustering: K-Means / K-Medians / K-Medoids
 - b. Hierarchical Clustering
 - c. Fuzzy Set Theory
 - d. Fuzzy Clustering: FCM
 - e. Probabilistic and Generative Clustering: Expectation – Maximization Algorithm
 - f. Density Based Clustering: DBSCAN
6. Association Rule Mining
 - a. Basic Terminology
 - b. *A priori* Algorithm
 - c. Applications in Operations Management
7. Data Understanding & Visualization
8. Data Preprocessing

- a. Outlier Determination and Handling
 - b. Imbalanced Data
 - c. Missing Data
 - d. Scales & Normalization of Data
9. Meta Heuristics
- a. Local Search Algorithm (i.e., Hill Climbing)
 - b. Extended LSA
 - c. Variable Neighborhood Search
 - d. Simulated Annealing
 - e. Tabu Search
 - f. Genetic Algorithms
 - g. Beam Search
 - h. Greedy Randomized Adaptive Search Procedure (GRASP)
10. Feature Subset Selection
- a. Filtering Methods
 - b. Wrappers
 - c. Embedded Techniques
11. Feature Extraction
- a. Principal Component Analysis
12. Neural Networks
- a. Perceptron
 - b. Activation Functions
 - c. Feed Forward
 - d. Back Propagation

Course Schedule and Disclaimer

The course schedule can be found in Schedule of IE 525 Spring 2024.xlsx 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.