

Course Objective

Data Science is the study of extracting information from data. Addressing the needs of the industry requires skills in data processing and data analytics spanning a wide area of subjects ranging from Statistics to Machine Learning. This course will introduce some of the basic concepts, techniques and tools that are required to solve problems widely seen in data analytics. The course will start with a short review on inferential Statistics and exploratory data analysis. The focus in the study of these subjects will be breadth, rather than depth, and practical examples will be used in applications of classification and regression techniques to a wide variety of problems in predictive data analytics.

Projected Outline

- **Introduction and Terminology**
- **Python Recap & Exploratory Analysis with Pandas and Matplotlib**
- **Linear Regression**
- **Scikit Learn – A Deeper Look**
- **Model Accuracy and Bias-Variance Trade-off**
- **Introduction to Classification: Logistic Regression**
- **Regularization and Model Tuning**
- **K-Nearest Neighbors**
- **Tree Based Methods**
- **Support Vector Machines (Optional)**
- **Natural Language Processing (Optional)**
- **An unsupervised approach: Dimensionality Reduction and PCA (Optional)**

Grading

- Labs : 25%
- Homework : 20%
- Quizzes (2) : 25%
- Course Project : 30%

Course Project

A problem will be submitted for solution. In the final week of the course you'll present your solution. Projects are required to be in groups of at least 2.

Requirements: Notebook and Presentation files + Presentation in class

Presentation Details

- *Problem Description*
- *Exploratory Analysis (20%)*
- *Data Preprocessing (20%)*
- *Feature Selection (20%)*
- *Model Tuning (20%)*
- *Results (20%)*
- *Conclusion*

Supplementary Books

- **Introduction to Statistical Learning**
 - James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.
 - Full Book Page:
 - <https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf>
- **Machine Learning Mastery** – Understand You Data, Create Accurate Models and Work Projects End-To-End, Jason Brownlee
 - <https://machinelearningmastery.com/machine-learning-with-python/>
- [Scikit-Learn Website](#)

Schedule

Class Hours:

Saturdays 13:00 pm

Tuesdays 19:00 pm

Contact

Altuğ Tanaltay

Office : *MDBF – L026 (08:30 – 17:30)*

Office Tel : *0(216) 4839561*

Email : *atanaltay@sabanciuniv.edu*