

Web Analytics and Mining

MET CS 688

Course Format (On Campus)

Reza Rawassizadeh

rezar@bu.edu

Office hours: by appointment

Course Description

In this course students learn web scrapping, crawling concepts, technologies and legal issues associated with them. Then, the course focus shifts to statistics required for unsupervised learning. The major part of this course focused on applying unsupervised learning algorithms to web data, including clustering and graph algorithms. Besides, students will get familiar with dimensionality reduction techniques in the context of web mining and web search algorithms. Students who attend this course should be familiar with both R and Python programming, but there is no need to have a robust machine learning or statics background.

Required Book

“Machine Learning and Artificial Intelligence: Concepts, Algorithms and Models” ISBN: 979-8992162103

Course Requirements

Students should be familiar with R and Python programming. Besides, they need to pass CS 555 and CS 544.

Class Policies

- 1) **Attendance & Absences** – Class attendance is not mandatory but highly recommended.
- 2) **Assignment Completion & Late Work** –About 40% to 70% of final grade is coming from assignment delivery. Late submission of homework is associated with a penalty of 10% grade reduction for any single day.
- 3) **Academic Conduct Code** – “Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions. Please take the time to review the Student Academic Conduct Code:

http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.html.

Grading Criteria

40% of the final grade is coming from assignments, 30% from final project delivery which is a scientific report about assignments and the last 30% are from final exam. Students who might require assistive grade could do a scientific presentation in the class and this can provide up to 10% additional credit on their final grade.

Class Meetings, Lectures & Assignments

Lectures, Readings, and Assignments subject to change, and will be announced in class as applicable within a reasonable time frame.

Date	Topic
Session 1	Web Analytics, Scraping, and Crawling
Session 2	Introduction to Machine Learning and Visualization
Session 3	Feature Engineering (numerical, categorical data), Word Embeddings and Language Models
Session 4	Clustering (similarity metrics, partition-based clustering, density-based clustering, hierarchical clustering, probabilistic clustering, big data clustering)
Session 5	Graph and Search Algorithms I
Session 6	Graph and Search Algorithms II