Boston University Metropolitan College

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 on statistics required for unsupervised learning. The major part of this course focused on applying unsupervised learning algorithms on web data, including association rule mining, sequence mining and clustering. 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 basic R and Python programming and there is no need to have a machine learning or statics background.

Books

There is no specific book required for this course, slides and in class presence are enough.

Class Policies

1) **Attendance & Absences** – Class attendance is not mandatory but highly recommended.
2) **Assignment Completion & Late Work** – About 40% 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	Assignments Due
Session 1	Web Analytics, Scraping, and Crawling	Six days after Session 1
Session 2	Introduction to Machine Learning and Visualization	Six days after Session 2
Session 3	Introduction to Statistics required for Web Mining I	NA
Session 4	Introduction to Statistics required for Web Mining II	Six days after Session 4
Session 5	Feature Engineering in the Context of Web Data	Six days after Session 5
Session 6	Sentiment Analysis and Theme Extraction	Six days after Session 6
Session 7	Clustering I (similarity metrics, partition based clustering, density based clustering)	NA
Session 8	Clustering II (hierarchical clustering, probabilistic clustering, big data clustering)	Six days after Session 8
Session 9	Association Rule Mining	NA
Session 10	Sequence Mining	Six days after Session 10
Session 11	Dimensionality Reduction Methods in the Context of Web Datasets	Six days after Session 11
Session 12	Web Search Frameworks, Page Rank, HIT	NA
Session 13	Project Presentations	NA
Session 14	Review	NA