

Web Mining and Graph Analytics

MET CS 688 A1

On Campus

| | | |
|--------------------|--------------------|--|
| Instructor | Parivash Hajiyani | hajiyani@bu.edu |
| Teaching Assistant | Keerthi Uppalapati | keer2003@bu.edu |

Office hours: by appointment

Course Description

The Web Mining and Graph Analytics course covers the areas of web mining, machine learning fundamentals, text mining, clustering, and graph analytics. This includes learning fundamentals of machine learning algorithms, how to evaluate algorithm performance, feature engineering, content extraction, sentiment analysis, distance metrics, fundamentals of clustering algorithms, how to evaluate clustering performance, and fundamentals of graph analysis algorithms. Laboratory Course. Prerequisites: MET CS 544, or MET CS 555 or equivalent knowledge, or instructor's consent.

Books

Machine Learning and Artificial Intelligence: Concepts, Algorithms and Models by Reza Rawassizadeh
https://www.amazon.com/dp/B0DV3WWMBD?ref=ppx_yo2ov_dt_b_fed_asin_title

Course References

- 1- Web Scraping with Python: Collecting More Data from the Modern Web
By Ryan Mitchell · 2018
<https://www.barnesandnoble.com/w/web-scraping-with-python-ryan-mitchell/1126320393>
- 2- Introduction to Machine Learning with Python
By Andreas Müller C. Sarah Guido · 2016
<https://www.barnesandnoble.com/w/introduction-to-machine-learning-with-python-andreas-c-m-ller/1140203548>

Courseware

Course material will be posted on Blackboard.

All the adjustments and changes will be announced through the course Blackboard page:
https://learn.bu.edu/ultra/courses/_105804_1/cl/outline and students are responsible to review the material and announcements.

Class Policies

- 1) **Attendance:** Mandatory
- 2) **Assignment Completion & Late Work** – 55% 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** – For students, the Code establishes an environment of integrity and professionalism that helps to assure each individual of receiving appropriate recognition for their

work. The ethical decisions that students face in an academic environment are similar to those they will encounter routinely in the professional world they will enter upon graduation or where they are currently employed. The Code allows faculty to conduct a fair and accurate evaluation of student performance and to maintain a supportive and just learning environment. ***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:

<https://www.bu.edu/academics/policies/academic-conduct-code/>

Grading Criteria

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.

| | Breakdown |
|--------------|-----------|
| Attendance | 5% |
| Midterm Exam | 15% |
| Final Exam | 25% |
| Assignments | 55% |

Class Meetings, Lectures & Assignments

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

| Date | Topic | Assignments Due |
|------------|---|--------------------------|
| Session 1 | Fundamentals of Machine Learning | NA |
| Session 2 | Web Scraping and Web Crawling | 2 weeks after session 2 |
| Session 3 | Review on Statistics required for Web Mining I | NA |
| Session 4 | Review on Statistics required for Web Mining II | 2 weeks after session 4 |
| Session 5 | Feature Engineering I | NA |
| Session 6 | Feature Engineering II | 2 weeks after session 6 |
| Session 7 | Language Models and Word Embedding | NA |
| Session 8 | Sentiment Analysis and Theme Extraction | 2 weeks after session 8 |
| Session 9 | Clustering I | NA |
| Session 10 | Clustering II | 2 weeks after session 10 |
| Session 11 | Graph and Search Algorithms I | NA |
| Session 12 | Graph and Search Algorithms II | NA |
| Session 13 | Dimensionality Reduction and Data Decomposition | NA |
| Session 14 | Review | NA |