

**Boston University Metropolitan College**  
COMPUTER SCIENCE DEPARTMENT



## **Web Analytics and Mining** **MET CS 688 A2**

### **Instructor:**

Zlatko Vasilkoski

email: [zlatko@bu.edu](mailto:zlatko@bu.edu)

Office hours: by appointment

**Class Time:** Thursdays from 6:00pm to 8:45pm,

Location: HAR 312

Semester Starts: January 20, 2022, and ends on April 28, 2022.

### **BU Community COVID-19 Public Health Policies**

All students returning to campus will be required to be [vaccinated against COVID-19](#), and upload information about their status (including applications for a medical or religious exemption or an extension) to the [Patient Connect](#) portal. In addition to the vaccine requirement, students must follow all other safety protocols, including the [face covering policy](#), and [screening](#), [contact tracing](#), and [testing](#) requirements. At the beginning of each class you will be asked to show a green [Healthway](#) compliance badge on your mobile device to the instructor, and wear your face mask over your mouth and nose at all times.

### **Course Description**

The Web Analytics and Mining course covers the areas of web analytics, text mining, web mining, and practical application domains. The web analytics part of the course studies the metrics of web sites, their content, user behavior, and reporting. The text mining module covers the analysis of text including content extraction, string matching, clustering, classification, and recommendation systems. The web mining module studies how web crawlers process and index the content of web sites, how search works, and how results are ranked. In addition, certain aspects of cloud computing will be illustrated, through an introduction to Internet of Things (IoT) analytics. Application areas in mining the social web and game metrics will be also investigated. Over the last two years, great advances have been made in the field of Natural Language Processing (NLP) and we will have an overview of the latest models such as Transformers (BERT), Reformers, etc. We will also illustrate the very latest NFT technology that gained great popularity since March of 2021.

Please note that this is a laboratory course in which in-class use of computers is required. The class is not scheduled in a lab with computers thus laptops are required. Also note that substantial programming experience with R is needed for this course. Some of the assignments involve creation of Shiny web apps that you may run on mobile devices and data analytics involving optimization of neural networks. In addition, the latest NLP tools are available in Python so some experience with it would be very helpful.

### **Required Course Prerequisites**

CS544 (Foundations of Analytics) or CS555 (Data Analysis and Visualization).

### **Course Grading Policy**

100–93.00	A
92.99–90.00	A–
89.99–87.00	B+
86.99–83.00	B
82.99–80.00	B–
79.99–77.00	C+
76.99–73.00	C
72.99–70.00	C–
69.99–60.00	D
Below 60.00	F

The course grade will be based on active class participation and quizzes/tests/lab projects (10%), assignments (20%), midterm exam (35%), and a term project (35%). Assignments are expected to be submitted by their respective due dates. Late submissions carry a penalty.

### **Course Topics**

#### **Module 1 - Web Analytics**

- Metrics
- Key performance indicators
- Identifying important pages and web site visibility

#### **Module 2 - Web Analytics, Visualization and Statistics**

- Using Machine Learning for Web Analytics
- Dimensions and Segmentation
- Flow visualization, navigating reports

### **Module 3 - Text Mining**

- Preprocessing and content extraction
- Searching and fuzzy string matching
- Clustering text
- Classification, categorization, and tagging
- Question answering systems

### **Module 4 - Web Mining**

- Web Crawlers, Indexing
- Searching, precision and recall
- Ranking
- IoT (Internet of Things)

### **Module 5 - Applications - Mining the Social Web**

- Twitter - trending topics, Reddit - Social Graph API
- LinkedIn - Clustering colleagues, Google - Document Similarity

### **Module 6 - Applications - Game Analytics**

- Game metrics, telemetry, and analytics
- Telemetry collection and tools
- Game data analysis and visualization
- Case studies

### **Reference Textbooks**

Social Analytics: Network and Text Methods with NodeXL and R, Shaila Miranda, 2018

Michael Beasley, Practical Web Analytics for User Experience: How Analytics can help you Understand your Users, Morgan Kaufmann, 2013.

Grant S. Ingersoll, Thomas S. Morton, and Andrew L. Farris, Taming Text: How to find, organize, and manipulate It, Manning Publications, 2013.

Matthew A. Russell, Mining the Social Web, 2nd Edition, O'Reilly, 2013

Magy Seif El-Nasr, Anders Drachen, Alessandro Canossa, eds., Game Analytics: Maximizing the Value of Player Data, Springer, 2013.

Barney L. Capehart, Ph.D., C.E.M., Michael R. Brambley, Ph.D., Automated Diagnostics and Analytics for Buildings, Fairmont Press, 2014.

**Course Web Site**

<http://learn.bu.edu>

**Student Conduct Code**

<http://www.bu.edu/met/for-students/met-policies-procedures-resources/academic-conduct-code/>