



BA820 – Unsupervised and Unstructured Machine Learning

Spring 2024

Course Administrative Details

- **instructor:** [Mohannad Elhamod](#) (Contact exclusively through [Piazza](#))
 - **office hours:** Wednesdays 4:00 pm – 5:00 pm/ HAR 546D or [Zoom](#).
- **TAs:** (Contact exclusively through [Piazza](#))
 - **office hours (all through Zoom):**
 1. Atharva: Thursday 16:00 – 17:00
 2. Gowtham: Tuesday 12.30 - 13.30
 3. Mugdha: Monday 17:00 - 18:00
- **Class time and place:**
 - A1: MW 8:00 am – 10:45 am / HAR 324
 - B1: MW 12:45 pm – 3:30 pm / HAR 324

The instructor reserves the right to update the syllabus at any time.

Course Description & Learning Goals

This course will cover the methods being applied to both unstructured and unlabeled datasets. Through a series of lectures and hands-on exercises, students will examine the techniques to unlock insights from data that appear to lack a known outcome. The goal of this course is to compare and contrast the application of various methods being applied today and provide the foundation to develop impactful insights from these datasets.

Upon completion of this course, you will have learned:

- How, and when, to apply Unsupervised Machine Learning techniques.
- How to work with text as a data source to enrich our analyses and machine learning models.
- How to apply the content in this course to real-world business challenges, with a focus on the ability to communicate and defend the results.

Diversity & Inclusion Statement

In this classroom, your value and dignity are recognized. We deeply respect your academic freedom and encourage you to express thoughts, pose challenging questions, and explore diverse viewpoints. We are dedicated to cultivating an atmosphere where everyone feels welcome and respected. Respectful dialogue is not just encouraged—it's essential. Critical thinking is at the heart of our learning process. By upholding these principles, we aim to create an inclusive and intellectually stimulating atmosphere that celebrates the pursuit of knowledge for everyone.

Course Materials and Logistics:

- **Learning platforms and tools:** The following tools will be used during this course:
 - **Datacamp** – Use the following [link](#) to associate your @bu.edu email with the class. Datacamp will host some of the reading materials as well as some assignments/tests. This is specific to this offering:
 - **Github Account and Github Desktop** - Desktop installation [here](#). We will use [Github Classrooms](#) for some assignments and tests. All slides and class resources will be found through this [Github repository](#).
 - In addition to the above, it is assumed that you have the ability to analyze data via **Python in Google Colab**.
- **Textbooks and supplemental material:** While there isn't a primary textbook for this course, the following are excellent resources when learning to programmatically analyze data using python. The links below are free online and can be used as supplemental references for this course and other offerings within this program.
 - [Hands-on Machine Learning with Python Implement Neural Network Solutions with Scikit-learn and PyTorch](#) (via BU Library)
 - [Explainable AI with Python](#) (via BU Library)
 - [Text Analytics with Python](#) (via BU Library)
 - [Natural Language Processing Recipes](#) (via BU Library)
 - [Pro Machine Learning Algorithms - Chapters 8,11-13](#) (via BU Library)
 - [Python Data Science Handbook](#) (Available online as an ebook)
- **Class format:** A typical session will begin with an informative lecture introducing new concepts. This will be followed by an interactive, hands-on segment designed to deepen understanding and allow for practical application of these concepts from both a technical and real-world perspective.

Course Policies

- Contact Policy

All communications for this course will exclusively be through [Piazza](#). This includes sending messages to your instructors privately or posting and replying to questions and discussion topics. When posting a question to your instructor/TAs, make sure it is posted to ALL of them. This helps getting your question answered promptly. Any communiques sent through Blackboard or direct email will most likely not receive attention. For a tutorial on how to use Piazza, please refer to this [video](#).

- Attendance Policy

Satisfactory class contributions require regular attendance (including arriving to class on time); preparation of all materials for every session; and active, quality participation in class discussions. Simply attending class, however, does not constitute a positive contribution to class and will not yield high class contribution scores.

The following table outlines the penalty for missing/arriving late to classes:

Number of unexcused absences or late arrivals	Penalty
2	-
3	3 points
4	6 points
6	An entire letter grade of final grade.
7 or more	Failing the course

Finally, for 'excused absences' (See policy [here](#)), the instructor will, where feasible, help arrange for missed graded components.

- Academic accommodations for students with special needs

In keeping with university policy, any student with a disability who needs or thinks they need academic accommodations must call the Office of Disability Services at 617-353-3658 or stop by 19 Deerfield Street to arrange a confidential appointment with a Disability Services staff member. Accommodation letters must be delivered to your instructor in a timely fashion (not later than two weeks before any major examination). While your instructor is happy to accommodate reasonable requests when possible, please note that accommodations may still not be delivered absent an official letter of accommodation.

- Academic Integrity Policy

- **Standard policy:** Please refer to the [university's general academic integrity policy](#) and [code of conduct](#). Unless specifically instructed to the contrary, these policies will be enforced.
- **Use of AI:** We live in interesting and disruptive times where technology is moving faster than our ability to grapple with it. Still, integrating the latest technological tools into your knowledge and skillset is more important now than ever. During this course, you will be allowed to use AI tools, **but only under specific conditions set by your instructor for each task. Unless explicitly instructed, the use of AI for a task is strictly forbidden.** If AI use is permitted, you must clearly disclose the prompts you used and properly credit the

AI tool(s) in APA Style. You may also be required to provide evidence of your prompts and the AI's responses. It is important to note that simply replicating AI-generated content without any intellectual engagement will be regarded as plagiarism. Non-compliance with these rules constitutes a violation of the honor code and will result in appropriate disciplinary actions.

- **Project Collaboration:** The ultimate goal of education is to help you prepare for success in real-life. Recognizing the critical role of teamwork and communication in professional environments, this class includes a project component. In this component, teams of 4-5 students will work together to produce results. However, it's crucial to **remember that each student is expected to fully grasp, discuss, defend, and take ownership of the entire project.** Therefore, all team members must actively engage in various aspects of their projects, including research, coding, documentation, and presentation. This ensures a comprehensive understanding and contribution from every member. Significant imbalances in the distribution of responsibilities or the type of work undertaken by team members will result in penalties.
 - **Citations and References:** For all submissions, you are required to provide citations and references to any articles, repositories, or other materials you use. Omission will lead to losing points.
- **Professional Conduct Policy**
- **Cellphones are prohibited** unless specifically allowed by the instructor for certain in-class activities.
 - **Place your name tent:** This helps your instructor learn your names. *It is also needed so your TA could record your participation in class.*
 - **Bring your laptop:** You will be required to use your laptops for in-class activities such as coding. While laptops and tablets are acceptable for notetaking, ensure they are not to cause disruptions or distractions. Also, don't forget your power cords.
 - **Food and drinks are not allowed**, except for water. A 15-minute break will be provided midway through the class.

Course Evaluation & Expectations

For details on Questrom's program-wide guidelines for grading MSBA courses, please refer to the following [link](#). If you have any questions about grades that you receive on particular assignments, you must raise them within a week of receiving your grade on that assignment. If you have particular grade-related considerations that you think are important, please raise these with your instructor as early as possible (during the first half of the semester at the latest!), so that your instructor can help you approach the course in a way that will help you achieve your best possible performance.

Grades will be determined according to the following table:

Grading Component	Percent of Course Grade
Midterm exam	20%
Class participation	10%

Individual Assignments	15%
Team project	25%
Final Exam	30%
Total	100%

- **Midterm (20%) and Final Exams (30%):** The two exams will be based on class notes, in-class discussions and exercises, and individual assignments outside of class. Anything discussed in class can be tested, unless explicitly excluded. Therefore, being present and participating in class and working through the individual assignments and DataCamp chapters is critical for success on the exams. The exam format will be mostly short-answer questions and problem sets aimed at assessing analytical reasoning and ability to produce insights using python.
- **Class Participation (10%):** Regularly engaging in discussion of the topics, asking questions that lead to better understanding of a concept by the class as a whole, clarifying concepts, and sharing professional experience about course topics constitute superior class participation and contribute to our collective learning. Simply attending class is not a pathway to full credit. Consideration for active participation on the forums will be included in this component of the grade. A tentative grade will be shared mid-semester as a form of feedback.
- **Individual Assignments (15%):** The individual assignments are intended to reinforce the fundamentals covered within the course while simultaneously building skills using tools commonly found in today's data science pipelines. Assignments are always due based on the posted date and time, even if you are unable to attend class that day. These assignments are graded on two dimensions: technical accuracy (e.g. the code functions properly and is consistent with the stated problem) and analytical reasoning and presentation of the results. All assignments are to be completed individually.
- **Team Project (25%):** In assigned teams of 4-5 students, collectively identify a dataset that can be used to solve a stated business problem. Using the techniques covered in class and any topic covered in the program, explore and analyze the data to answer the defined problem and present your work and findings in a 10–15–minute presentation on the 12th session. Attendance is mandatory. The following are the detailed project steps/deliverables:
 - Students will be allowed to self-select their groups.
 - **Deliverable 1, A 2-page project proposal:** The proposal should summarize (a) the problem that motivates the analysis you will conduct, (b) the data set you plan to use, and if possible (c) your proposed analysis methodology.
 - **Deliverable 2, A mid-project report, no longer than 2 pages:** This report should discuss initial results after completing your exploratory analysis and preliminary analysis. For this deliverable, you may provide an additional document with tables/figures.
 - **Deliverable 3:**
 - A final write-up discussing (a) the problem, (b) the data, (c) the methods applied and (d) your analytical findings, conclusions and recommendations as they relate to your business problem.
 - A 10–15–minute presentation on the content and structure, like the write-up.
 - A reference or appendix of your data sets.

BA820 - Unsupervised and Unstructured Machine Learning

- The scripts/project that you used to complete your work. This can include python scripts, notebooks, Streamlit apps, links to Github Repositories, etc.

Your final presentation will include a peer-evaluation component. The stakeholders are students on another team; they will ask you questions about your project, your process, and results. All students will submit an evaluation form to the instructor based on your presentation, work, and responses to questions. The final project grade will be based on their evaluation, the instructor's evaluation, and an assessment from your teammates. Peer evaluation will count as 30% of the overall project grade, with the remaining 70% coming from the instructor's assessment.

- **Late Submission:** 10% will be deducted from a deliverable grade for each midnight that passes following the deadline indicated in the course schedule, until a deduction of 30% has been reached, beyond which a grade of 0 will be assigned.

Other Logistics

- **Blackboard Usage:** Though several learning platforms and tools are used in this class, Blackboard will act as the hub from where all information can be found. All deliverables and their deadlines will be posted through Blackboard. It is essential you pay close attention to announcements, which contain critical information, such as schedule updates, assignment feedback, and other valuable discussions. Also, your instructor will regularly update the class website with fresh materials and assignments. And while a weekly announcement will generally be sent as a reminder, it is still your responsibility to setup your email alerts appropriately and regularly for any updates in the schedule, assignments, or reading materials.

Course Schedule

Note that while the following table provides a holistic overview of the course's schedule, it is only meant to give general guidance. The exact dates, topics, readings, and deliverables will be posted on Blackboard. Whenever there is a conflict between the syllabus and Blackboard, Blackboard is correct.

<i>Session</i>	<i>Date</i>	<i>Readings/Podcasts/Datacamp</i>	<i>Learning Goals</i>	<i>Deliverable</i>
1	1/22	Market Basket Analysis in Python	Association Rule Mining	
2	1/24	Cluster Analysis in Python Chapters 1 and 2	Distance Metrics & Hierarchical Clustering	
3	1/29	Cluster Analysis in Python Chapter 3	K-Means Clustering	Team Project Formation DUE

BA820 - Unsupervised and Unstructured Machine Learning

4	1/31	Unsupervised Learning in Python - Chapter 3	Clustering Wrap Up Dimension Reduction 1	Assignment 1 Released Team Project Deliverable 1 Released
5	2/5	Unsupervised Learning in Python - Chapter 2	Dimension Reduction 2 / Putting it all together	
6	2/7			Exam 1
7	2/12	Intro to Natural Language Processing in Python - Chapter 1	Working with Text	Team Project Deliverable 1 DUE Team Project Deliverable 2 Released
8	2/14	Feature Engineering for NLP - Chapter 4	Document Clustering	Assignment 1 DUE
-	2/19	-	NO CLASS	Assignment 2 Released
9	2/21	Intro to Natural Language Processing in Python - Chapter 3	Sentiment Analysis/Classification	
10	2/26	Feature Engineering for NLP - Chapter 2 <i>Identifying New Materials with NLP - TWiML Podcast</i> <i>Named Entity Recognition - Data Skeptic Podcast</i>	Intent Classification and NER	Team Project Deliverable 2 DUE Team Project Deliverable 3 Released
11	2/28	Building Enterprise Grade AI Assistants in Production Topic Modeling for non-ML Folks The Voicebot Podcast	NLP Wrap-up Putting it all together	
12	3/4			Assignment 2 Due Final Projects Presentation
13	3/6			Exam 2

--	--	--	--	--