**Boston University** Questrom School of Business

# BA810: Supervised Machine Learning


| | |
|---|---|
| Instructor: | Nachiketa Sahoo |
| Office Phone: | 617-353-6145 |
| e-mail: | nachi@bu.edu |
| Office Hours: | Mon, Wed 1pm – 2pm<br>HAR639A and Zoom link. |


Teaching Assistants and office hours (in person in **HAR427** and respective zoom links):

1. Barrett Ratzlaff (baratz00@bu.edu), Thu 5–6pm, zoom link
2. Tharfeed Ahmed Unus (tharfeed@bu.edu), Tue 1–2pm, zoom link
3. Ran Zhang (ran0925@bu.edu), Fri 3–4pm, zoom link

## COURSE DESCRIPTION & LEARNING GOALS

This course will teach students how to perform hands-on analytics on real-world datasets using supervised machine learning techniques. Students will analyze data using Python programming language, derive actionable recommendation from data, and present their findings. The goal of the course is to create an understanding of popular supervised machine learning techniques and the types of problems they can be applied to.

## PREREQUISITES

- MSBA Python and Statistics bootcamp
- BA780

## COURSE STRUCTURE & PEDAGOGY

The general structure of each class will be as follows:

- Lecture and discussion (first half). Every lecture will start with a recap of the previous lecture: if you have questions from the prior lecture (or any other prior material), this is a great time to ask.
  - *Close your laptops during this time.*
- 10 min break
- Lab, coding session, results discussion (second half of class)

*Cell phones should be put away during the entire class.*

The lab part of each class will apply lecture concepts to real-world datasets using Python. Bring your laptop to each class: they are necessary to do the lab exercises. *Use your laptop only for in class exercises.*

## DIVERSITY AND INCLUSIONS STATEMENT

In developing this course, I have aimed to be thoughtful about how identity and culture impact the course content. I invite you to share your personal experiences and perspective related to the course content. If there are topics or conversations that you feel would benefit from incorporation of social context, a differing perspective, or Questrom's Office of Diversity & Inclusion, please inform me and I will explore resources and opportunities for us to engage a wide variety of perspectives in our classroom.

## COURSE MATERIALS

- ISL: "An Introduction to Statistical Learning with Applications in Python" by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. The textbook is available for free from the authors' website: https://www.statlearning.com/ .
- HOML: "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow." (2022, 3e) by Géron, A. This book emphasizes the practical side of machine learning. A digital copy of the book is available through BU library. You can use this link to access.
- CPSC: Chelsea Pelleriti's video lectures (https://www.youtube.com/@ChelseaPelleriti/playlists) as supplementary material . We will draw mostly from the course CPSC392 and a few from CPSC393.

See the course schedule at the end of this document to see the sections of the books you need to read before each class. Therein, you will also find Datacamp chapters that need to be completed for each class as well links to optional supplementary material that may be helpful to understand the material discussed in each lecture.

Other course material (slides, sample exams, instructions, etc.) will be made available through this google drive folder. You need to be logged into the BU Google account to access it. We'll use the course blackboard to collect most deliverables.

### Software

We'll use Google Colab to write and run our code. Review the Overview of Colaboratory Features before the first class, if you need a refresher. We'll use scikit-learn and other related Python packages to analyze datasets.

## COURSE POLICIES

### Attendance

Attendance is mandatory except as permitted by the Boston University rules. Unexcused absences will hurt your grade.

### 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 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 me in a timely fashion (within two weeks of the date on the letter and not later than two weeks before any major examination). Please note that accommodations will not be made absent an official letter of accommodation.

### Mental Health and Wellness

Life at college can get complicated and it is easy to feel overwhelmed, lost, anxious, or depressed. If you find yourself struggling with your mental or physical health this semester, please feel free to approach me. I will try to be flexible and accommodating, within reason. But I am not a professional therapist and there is no shame in getting help. Help for managing stress and your mental wellbeing can be found at Student Health Services. There you can find short term therapy, groups, and workshops, 24/7 on-call service (617-353-3569), referrals, and more resources. If you are feeling stressed and having trouble making choices around alcohol consumption the Collegiate Recovery Program may offer help.

**Financial Insecurity**

We learn as whole people. It can be challenging to do your best in school when you are worried about meeting basic needs like safe shelter, sleep, and nutrition. If financial insecurity is an obstacle to learning for you I urge you to contact Terrier Meal Share, Boston University Financial Assistance, or the Dean of Students Office.

**Academic Integrity**

My goal is to create a community in BA810 where you will feel invested and included. The importance of honesty and integrity in this learning community cannot be overemphasized. Every member of the Boston University Questrom School of Business is responsible for creating an ethical environment. I will do my best every day to create that in our class, but it can't happen without your commitment too. I expect you to abide by the Academic Conduct Code. It is easy to do: follow all assignment and exam rules. If you have a question about the rules of an assignment, (Can I collaborate on this assignment? Can I use my phone to check the time during an exam? Can I Use GenAI Tool on this?) ask me.

Any student caught cheating or submitting work that is not their own will be referred to the Academic Committee for investigation which may result in a variety of penalties including removal from the program. They most likely will be ashamed of themselves and embarrassed for their lapse in judgment for years to come.

Please see BU's Academic Conduct Code for more information. And always remember: if you have questions or concerns about an assignment, deadline, reading, exam, or anything related, just ask me.

**Classroom Conduct**

- **Professionalism**: Students are expected to follow the Boston University's Student Codes of Conduct. They are expected to conduct with a sense of respect and professionalism and can expect the same treatment. A part of our learning goal is what is expected in a workplace, and we try to enforce those standards throughout the program.
- **Punctuality**: Students are expected to arrive at the classes and scheduled meetings on time. Do return to class promptly after breaks.
- **Name Tents:** Students are expected to keep the name tents in front of them during the class. This is necessary for getting participation grades. This also helps peers and faculty know them faster.
- **Participation:** Students are expected and strongly encouraged to ask questions and get involved in class discussions. This helps everyone get more out of the class.
- **Cell Phone:** Students cannot use their cell phone during class or exams unless specifically instructed. If you need to take an urgent call, simply leave the classroom, and return as soon as possible.
- **Activities Unrelated to Class:** Activities that are unrelated to the class are not allowed during the class. These include, but are not limited to, using social media, news sites, online video sites, gaming, checking/writing emails, unless asked by the instructor for the purpose of the class.
- **Absence Policy -** If you need to miss a class, notify your instructor before the class. Please review the recording and slides for the missed classes as soon as you can and contact the instructor if you need help on the missed material.

**Sexual Misconduct/Title IX Policy**

The Questrom School of Business is committed to fostering a safe learning environment for all members of its community and preventing sexual misconduct. All forms of sexual misconduct, including rape, acquaintance rape, sexual assault, domestic and dating violence, stalking, and sexual harassment are violations of Boston University's policies, whether they happen on campus or off campus. Title IX of the

Education Amendments of 1972 is a federal civil rights law that prohibits sex-based discrimination in federally funded education programs and activities. This law makes it clear that violence and harassment based on sex and gender is a Civil Rights offense subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, etc. If you or someone you know has been harassed or assaulted, you can find the appropriate resources at https://www.bu.edu/shs/survivor-support/ .

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility in my role as a faculty member. It is my goal that you feel able to share information related to your life experiences in this class. I will seek to keep the information you share private to the greatest extent possible. However, I am required to share information regarding sexual misconduct with the University. Students may speak to someone confidentially by contacting Survivor Advocacy Response & Prevention Center.

**Class Recording Policy**

All class sessions will be recorded for the benefit of registered students Recorded sessions will be made available to registered students ONLY via their password-protected course accounts. Students may not share such sessions with anyone not registered in the course and may certainly not post them in a public platform. Students have the right to opt-out of being part of the class recording. Please contact me to discuss options for attending the course in such cases.

**Providing Feedback on the Course**

I'll collect anonymous feedback on the course from the students after first two weeks. You can provide such feedback on one or some classes anytime using this feedback form — I'd like to hear from you regarding how the class is going and what I can do to improve your learning.

**COURSE DELIVERABLES AND GRADING**

You will be graded according to the following criteria:

| | |
|---|---|
| Attendance and Class participation: | 15% |
| Individual assignments: 2 × 10% = | 20% |
| Datacamp assignments: | 15% |
| Team project: 5% (proposal) + 15% (final) = | 20% |
| Final Exam: | 30% |
| Total: | 100% |

**Class Participation**

Satisfactory class contributions require attending every session, preparing for every session by doing the Datacamp assignments and readings before class; and active, quality verbal contribution to in class discussions. Simply attending class, however, does not constitute a positive contribution to class and will not yield participation scores. I may call upon any student to answer questions during class. Please bring your name tent card to class so that I can accurately assign participation points.

We will grade participation in each class on the following rubric:

| Points | Type of contribution | Note |
|---|---|---|
| **1: Good** | Correct answer, good question, a good attempt, thoughtful question/answer that shows good engagement with material, adding helpful insights to others' comments | These are the expected contributions in a typical class. |
| **2: Outstanding** | Outstanding question/answer that creates a new learning opportunity for the class, making novel and useful connection between concepts | These are exceptional contributions. Typically, no more than 2-3 such comments are seen during a class. |
| **0: Minimal** | Comments lacking depth, showing a lack of preparation, repeating comments by others, asking to repeat or re-explain an idea. | While some questions are fair (e.g., clarifying a point), participation points are not expected. If a question is not of general interest, I'll defer it to one-on-one meetings. |
| **-1: per cellphone/ laptop usage policy violation** | Use of cell phone in class, use of laptop for purposes other than class work. | These hurt your learning and distracts others around you. |

To promote quality over quantity, multiple "Good" contributions does not make up for lack of "Outstanding" contribution. Neither will multiple "Outstanding" contribution get more than two points in a single class.

In sum, come to the class prepared, stay focused on class, and listen actively. Contribute value and give floor to others to improve collective learning.

**Individual Assignment**

The homework assignments need to be completed individually and submitted through blackboard by 11pm on the day they are due. Submit a PDF of the notebook with answer, which includes a link to the notebook that can be accessed if needed to verify code.

It is OK to ask your classmates for help, but you must disclose them (no penalty). It is not OK to copy someone else's (including those created by a generative AI tool) code into submission. All code and answers that you submit must be produced and written by you.

**Datacamp Assignment**

You need to complete certain Datacamp assignments before most classes --- see course schedule for details. These are due at 9am (i.e., before the class) on the day it is due (you may find it efficient to complete them well before the due time in batches that include related chapters). No late submission will be accepted for these assignments. These assignments can be accessed at Datacamp.com and need to be completed on Datacamp after joining the group set up for the class on the site. Invitation for the same will be sent out before the class starts. Nothing needs to be submitted from them on blackboard.

**Team Project**

You will work on projects in teams of four students. You will select teams during the first class. I will assign anyone unassigned at the end of first class to a team.

The project involves acquiring a dataset, performing a descriptive analysis on the dataset, and formulating & solving a prediction problem. There are two key steps. First, submit a two-page project proposal (through blackboard). Do meet with the professor by the date on the class schedule before submitting project phase 1. Ultimately, it is up to the instructor to approve the project and make recommendations.

On the last day of class, you will present your project. Submit the presentation slides and analysis notebook through blackboard by 9am on the day of the presentation.

Please see the "Project/ProjectInstructions.pdf" document on the course site for important details.

**Exam**

The final exam will be an offline closed-books-and-notes in-class exam. It'll consist of small coding and short answer questions. You are allowed to bring only one sheet of notes (both sided ok). No electronic device (laptop/tablet/phone, etc.) are allowed during the exam.

**Use of Generative AI and Other Assistance**

You may use a generative AI tool such as ChatGPT, BARD, or GitHub Copilot *to learn* how to solve a specific problem you are facing in the homework and the project. However,

1. *You cannot copy code generated by such a tool: you must write your own code. Anything you submit must be written by you.*
2. You must disclose in an appendix how you used such a tool.
3. You are ultimately responsible for the correctness of what you produce. Generative AI is known to make mistakes. Therefore, you should carefully consider what is generated by these tools and assess its correctness before adopting them.

Submitting code or text written by someone else, including generative AI, is not allowed.

**Grade Distribution**

Keeping with the voluntary grade distribution guidelines for MSBA core courses:

- Up to 40% of grades in the class will be A or A-
- Grades below A- will be assigned as earned

If you have any questions about grades that you receive on assignments, raise them within one week of receiving your grade on that assignment. Unless we have made computational errors, we will be unable to alter grades after final grades have been determined. If you have grade-related considerations that you think are important, please raise these with me as early as possible (during the first half of the semester at the latest!), so that I can help you approach the course in a way that will maximize your performance.

**COURSE SCHEDULE***

| # | Dates in 2025 | Topics, Readings | Datacamp | Deliverable |
|---|---|---|---|---|
| 1 | Mon, Oct 20 | Intro, types of ML problems, evaluation<br>R: ISL 2.1; HOML 1<br>S: CPSC392 Lec 3 | 1.1, 1.2, 2.1 | Team composition |
| 2 | Wed, Oct 22 | Regression, bias-variance<br>R: ISL 2.2, 3.1, 3.2<br>S: CPSC392 Lec 6, 7, 8 | 2.2, 3.2, 3.3 | |
| 3 | Mon, Oct 27 | Classification, evaluation metrics<br>R: ISL 4.1, 4.2, 4.3; HOML 3<br>S: CPSC392 Lec 9, 10, 14 | 4.1, 4.2, 4.3 | |
| 4 | Wed, Oct 29 | Cross validation, Regularization<br>R: ISL 5.1, 6.2 | 1.3 | HW1; meet prof/TA for project phase 1 |
| 5 | Mon, Nov 3 | Bootstrap, Model selection and tuning<br>R: ISL 5.2 | 5.1, 5.2, 5.3 | Project phase 1 |
| 6 | Wed, Nov 5 | Scikit-Learn, End-to-end example<br>*R: HOML 2 (particularly important)* | 1.4 | TLA feedback*** |
| 7 | Mon, Nov 10 | Advanced model selection (AutoML) | 5.4 | |
| 8 | Wed, Nov 12 | Feature selection<br>R: ISL 6.1, 6.4 | | HW2; |
| 9 | Mon, Nov 17 | Support Vector Machines and kernels<br>R: ISL 9.1, 9.2, 9.3; HOML 5<br>S: CPSC393 Lec 3, 4 | 4.4 | |
| 10 | Wed, Nov 19 | Decision Trees<br>R: ISL 8.1; HOML 6<br>S: CPSC392 Lec 11 | 6.1, 6.2, 6.3 | |
| 11 | Mon, Nov 24** | Ensembles<br>R: ISL 8.2; HOML 7 | 6.4, 6.5, 7.4 | |
| 12 | Mon, Dec 1 | Managing imbalanced data<br>R: Netflix case (in Slides folder) | | |
| 13 | Wed, Dec 3 | Review for exam | | |
| 14 | Mon, Dec 8 | Final Project Presentation | | Slides, notebook |
| 15 | Wed, Dec 10 | Final Project Presentation | | Slides, notebook, and TLA eval *** |
| 16 | Fri, Dec 12 | Exam: A1:1–3:30pm, B1:4pm–6:30pm.<br>HAR222 and HAR224 for both sections. | | |

* The schedule is tentative.

** Monday before Thanksgiving. It'll be a standard in-person class with no option to attend virtually.

*** Team Learning Assessments at teamlearning.bu.edu are due at the following times:

1. Mid-course *non-anonymous team feedback* due on Nov 5, 12noon.
2. End-of-course *anonymous team evaluation* due on Dec 10, 12noon.

R: Required before class.

S: Supplemental, can be read/watched before or shortly after class.

**DATACAMP MODULES**

1. Supervised Learning with scikit-learn
   1.1. Classification
   1.2. Regression
   1.3. Fine-tuning
   1.4. Preprocessing and pipelines
2. Introduction to Regression with statsmodels in Python

2.1. Simple Linear Regression Modeling
2.2. Predictions and model objects
2.3. Assessing model fit
2.4. Simple Logistic Regression Modeling

3. Intermediate Regression with statsmodels in Python
   3.1. Parallel Slopes
   3.2. Interactions
   3.3. Multiple Linear Regression
   3.4. Multiple Logistic Regression
4. Linear Classifiers in Python
   4.1. Applying logistic regression and SVM
   4.2. Loss functions
   4.3. Logistic regression
   4.4. Support Vector Machines
5. Hyperparameter Tuning in Python
   5.1. Hyperparameters and Parameters
   5.2. Grid search
   5.3. Random Search
   5.4. Informed Search
6. Machine Learning with Tree-Based Models in Python
   6.1. Classification and Regression Trees
   6.2. The Bias-Variance Tradeoff
   6.3. Bagging and Random Forests
   6.4. Boosting
   6.5. Model Tuning
7. Ensemble Methods in Python
   7.1. Combining Multiple Models
   7.2. Bagging
   7.3. Boosting
   7.4. Stacking
8. Extreme Gradient Boosting with XGBoost
   8.1. Classification with XGBoost
   8.2. Regression with XGBoost
   8.3. Fine-tuning your XGBoost model
   8.4. Using XGBoost in pipelines

The greyed-out chapters are optional.

**CHELSEA PELLERITI'S VIDEO LECTURES**

From CPSC392:

Lecture 3: All The Stuff You Need To Know (Math) Skip section on eigen vectors.
Lecture 6: Linear Regression I
Lecture 7: Linear Regression II
Lecture 8: Linear Regression III (Bias Variance Tradeoff)
Lecture 9: Logistic Regression I
Lecture 10: Logistic Regression II
Lecture 11: Tree Based Models
Lecture 14: Naive Bayes and K-Nearest Neighbors

From CPSC393:

Lecture 3: Support Vector Machines I
Lecture 4: Support Vector Machines II

Updated on: November 17, 2025