

**Syllabus**

**School of Professional Studies**

**DATA 622 Introduction to Machine Learning**

**Instructor Name**: Sabrina Khan

**Instructor Email Address:** [sabrina.khan@live.com](mailto:sabrina.khan@live.com) (preferred)

[sabrina.khan@sps.cuny.edu](mailto:sabrina.khan@sps.cuny.edu)

**Degree Program:** M.S. in Data Science

**Credits**: 3 graduate credits

**Prerequisites:** 605, 606, 621

**Type of Course**: Elective

**Meetup:** Every other week, on Wednesdays at 630pm EST, on Zoom. Attendance is optional. We start Feb 10, 2021.

**Office hours**: Tuesdays and Wednesdays 12pm -1pm EST, on zoom. Info available in the Instructor Bio & Contact info folder in BB

Feel free to reach out to get in touch via email and I will accommodate your schedule. My cell number is 929-245-0383, which may be easier to use if you want to text.

**Course Description:**

This course aims to develop basic understanding of foundational machine learning techniques. This course will aim to have a smooth transition from statistical modeling methods to machine learning methods. In this course, we emphasize on computational techniques. We will primarily use R as our programming environment. All assignments are expected to be to submitted in as R-Markdown documents. However, if you are more comfortable using Python, please let me know and we can make some adjustments.

**Learning Outcomes:**

1. Develop deep understanding of introductory machine learning algorithms
2. Prepare datasets for machine learning algorithms and conduct modeling exercises on given datasets.
3. Develop understanding to be able to identify problems that require supervised or unsupervised methods.
4. Develop proficiency of technical terminology expected of a Data Science practitioner.

**Letter Grade distribution:**

Please see: <https://sps.cuny.edu/about/policies/academic-and-student-policies/grading-policies-graduate>

**Late policy:**

Generally, late assignment is not accepted. If your discussion entries are late, there will be a 20% penalty.

**Assignments and Grading:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assignments** | **Possible score** | **Total points** | **% of the final grade** |
| Homework #1 and #2 assignments | 100 each | 100 x 2 = 200 | 10% each; 20% in total |
| Homework #3 and #4 assignments  (must done in groups) | 200 each | 100 x 2 = 400 | 20% each; 40% in total |
| 1 project  (must be done in group) | 250 | 250 | 25% |
| Weekly discussion:  Each week, you will enter your submission on the discussion prompt AND will respond to two other entries by your peers | 10 each | 10 x 15 = 150 | 1% each; 15% in total |
|  | Total | 1000 | 100% |

**Tentative Schedule (subject to change):**

| **Week** | **Week of** | **Topics** | **Key Task** | **Due on** |
| --- | --- | --- | --- | --- |
| 1 | 29-Jan | Introduction to 622, Intro to Machine Learning | Discussion 1 | 5-Feb |
| 2 | 5-Feb | Review week: Linear and Logistic Regressions | Discussion 2 | 12-Feb |
| 3 | 12-Feb | Classification: Naive Bayes, Linear and Quadratic Discriminant Analysis, k-Nearest Neighbors | Discussion 3 **Homework # 1** | 19-Feb |
| 4 | 19-Feb | Classification: Naive Bayes, Linear and Quadratic Discriminant Analysis, kNN (continued) | Discussion 4 | 26-Feb |
| 5 | 26-Feb | Tree based methods: Decision Trees | Discussion 5 | 5-Mar |
| 6 | 5-Mar | Tree based Methods: Bagging, Random Forests, Boosting Model Selection | Discussion 6 **Homework #2** | 12-Mar |
| 7 | 12-Mar | Tree based Methods: Bagging, Random Forests, Boosting Model Selection (continued) | Discussion 7 | 19-Mar |
| 8 | 19-Mar | Tree based Methods: Bagging, Random Forests, Boosting Model Selection (continued) | Discussion 8 | 26-Mar |
| 9 | 26-Mar | **Spring Break** |  | 4-Apr |
| 10 | 5-Apr | Support Vector Machines | Discussion 10 **Homework #3** | 9-Apr |
| 11 | 9-Apr | Unsupervised Learning: Clustering | Discussion 11 | 16-Apr |
| 12 | 16-Apr | Unsupervised Learning: Clustering… continued | Discussion 12 | 23-Apr |
| 13 | 23-Apr | Unsupervised Learning: PCA | Discussion 13  **Homework #4** | 30-Apr |
| 14 | 30-Apr | Resampling Methods & Model Selection | Discussion 14 | 7-May |
| 15 | 7-May | Resampling Methods & Model Selection (continued) | Discussion 15 | 14-May |
| 16 | 14-May | No meet up/optional | **Final Project** | 23-May |

**Required Texts and Materials**:

TITLE: An Introduction to Statistical Learning

AUTHORS: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

URL: <https://www.statlearning.com/> (booksite, pdf, data, R etc.)

<https://www.ime.unicamp.br/~dias/Intoduction%20to%20Statistical%20Learning.pdf> (Text we are following)

<https://www.dataschool.io/15-hours-of-expert-machine-learning-videos/> (Lecture videos and slides)

TITLE: Elements of Statistical Learning

AUTHORS: Jerome Friedman, Trevor Hastie, Robert Tibshirani

URL: <https://web.stanford.edu/~hastie/ElemStatLearn//printings/ESLII_print10.pdf>

(Text we are following)

More resources will be shared throughout the semester.

**Accessibility and Accommodations**

The CUNY School of Professional Studies is committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. For more information, please see:

[Disability Services on the CUNY SPS Website.](https://sps.cuny.edu/student-services/disability-services)

**Online Etiquette and Anti-Harassment Policy**

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University’s policies. Please see: [“Netiquette in an Online Academic Setting: A Guide for CUNY School of Professional Studies Students.”](http://catalog.sps.cuny.edu/content.php?catoid=2&navoid=205)

**Academic Integrity**

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see:

[Academic Integrity on the CUNY SPS Website.](https://sps.cuny.edu/about/dean/policies/academic-and-student-policies/academic-integrity)

**Student Support Services**

If you need any additional help, please visit [Student Support Services](https://sps.cuny.edu/student-services).