Machine Learning in Financial Engineering (FIN-UY 4903)

▼ Syllabus

Instructor Information

- Instructor: Professor Derek Snow
- Office: One MetroTech Center, 19th Floor
- d.snow@nyu.edu

Course Information

- Course Description: This course will introduce machine learning methods used in the world's largest hedge funds, banks, and other financial institutions.
- Co-requisite or prerequisite:
  - Proficiency in a programming language (like Python)
  - Familiarity with linear algebra and probability

Course Overview and Goals

Students will learn modern developments in machine learning and how it applies to finance. This course will give students the necessary vocabulary and technical expertise to participate in modern financial machine learning projects. The application of machine learning to asset management is highlighted throughout the course, i.e., trading strategies, weight optimization, and risk management.

Upon Completion of this Course, students will be able to:

- Understand the costs and benefits of **machine learning** and learn how to use it as part of a coding competition.
- Know that machine learning is a **process** that comprises (1) Preprocessing, (2) Model Development, and (3) Explanation steps.

- Preprocessing:
    - Through practical examples, learn the importance of **mismatch** preprocessing methods in developing models.
    - Learn why **data augmentation** methods like feature generation, label generation, and synthetic data generation is at the heart of every machine learning problem.
    - Understand the importance of **data selection** techniques including feature selection and row sampling.
- Model Development:
    - Understand what type of models you have in your toolbox and how to **select** and compare them.
    - Learn how to **optimize** your models through cross-validation for hyperparameter tuning and architecture learning.
    - Know the different performance and robustness metrics required when displaying your **results** and optimizing your models.
- Explanations:
    - Assess the various **feature importance** methods that can be used to understand the prediction models.
    - Know what types of **explanatory example** methods can be used for financial problems.
    - Finally review the main **explanatory visualization** methods that can help you narrate the outcome of your predictions.
- Learn about important **linear models** that underpin the foundation of financial modelling like OLS, LASSO, and Ridge Regressions.
- Explore **ensemble models** and learn how they are being used in large financial institutions as “grey-box models”.
- After we have dealt with some of the standard supervised learning models, we will give a quick overview of the following:
    - Learn the different types of **unsupervised learning** methods and their use cases.

    - See how **deep learning** has provided us a new paradigm to perform machine learning.
    - Know the importance of unstructured data, and how **natural language processing** can aid our prediction exercise.
    - Contend with the latest developments in **reinforcement learning** and understand how it is applicable in finance.
- Learn how these methods can be applied within an **asset management** framework that comprises trading strategies, weight optimization, and risk management.
- Finally if we have time, seek to understand some of the main future developments happening in this space, including **causal models**, **probabilistic models**, **graph neural networks**, **hybrid models**, and **automated machine learning**.

Course Requirements

Assignments and Grading

(1) The **first project** is the creation of a 10-slide presentation deck to convince a discretionary portfolio manager to use machine learning as part of their investment decision making process; you also have the choice of submitting a coding assignment instead. **(100 points).**

(2) The **next project** is the development and comparison of six different machine learning regression models as applied to an insurance modelling problem. You can use you preferred programming language. **(100 points).**

(3) The **third project** is writing a new blog post on any topic in financial machine learning, this can be technical or descriptive **(100 points).**

(4) The **fourth project** is the development of a notebook (code + explanation) that successfully engineers 10 unique types of features, two for each type of feature engineering: transforming, interacting, mapping, extracting, and synthesizing. **(100 points).**

(5) The rest of your points will come from **three takeaway homework** that would be made available to you after most weeks. **(90 points).**

(6) The last 10 points will come from participation during class like asking questions as well as asking questions on the forum **(10 Points)**.

> Your final grade will be out of **500 points**

👩🏽‍💻 **All these projects will be assessed the same way, I will look at every project from an employer's point of view.**

A well designed and written presentations, notebooks, code, blog posts, and summarized academic reports will show your future employer that you (1) understand the advantages and disadvantages of machine learning, (2) that you can write technically and/or descriptively about as specific topic in machine learning and finance, and (3) that you can read from, understand, and implement academic literature.

Tests & Quizzes

There will be no graded tests or surprise quizzes, you will be assessed during four independent projects and homework assignments for the quality of your coding skills, for the professionalism in your written work, and for participating in class discussions.

Rules for submission

Submissions and grading is performed on Brightspace.

Course Schedule

The course has three parts, (1) the first part will help us understand what machine learning is through the lens of classification and regression models as applied to cross-sectional and panel datasets, (2) the second part will investigate the step-by-step machine learning process all the way to production, (3) the third section will investigate additional machine learning innovations like unsupervised learning, deep learning, natural language processing, and reinforcement learning.

Topics and Assignments

# Week	Aa Topics
1	Introduction: History, Comparisons, Definitions, Applications in Finance, Linear regression
2	Regression Models: Linear regression, Ensemble Models, Time Series
3	Regression Models: Linear regression, Ensemble Models, Time Series
4	Classification Models: Linear, Ensemble, Time Series
5	Core Concepts: overfitting, regularisation, loss functions, feature engineering, exploratory data analysis, data munging.
6	Preprocessing: mismatch processing (type, length, missing, frequency, dimension), data augmentation (transforming, interacting, mapping, synthesising, extracting), feature selection (linear, non-linear, causal, conditional)
7	Model Development: specification (handcrafted, supervised, unsupervised, reinforcement learning, hybrid model), optimisation (hyperparameter tuning, cross validation), results (accuracy, robustness, performance)
8	Explanations: importance (permutation, conditional, relative), examples (prototypical, counterfactual, contrastive), visualisations (partial dependence, individual expectations, accumulated effects)
9	Unsupervised Learning: clustering and dimensionality reduction
10	Deep learning: classification, regression, and generative models
11	Deep learning: classification, regression, and generative models
12	Natural Language Processing: introduction + investigating all modern techniques
13	Reinforcement Learning: introduction + investigating all modern techniques
14	Asset Management: trading strategies (asset pricing, ranking), weight optimisation (ESG investing, unsupervised methods), risk management (recession prediction, regime switching, bet sizing)

Optional Readings & Text

You are not expected to read these, they are for your own personal development.

- Machine Learning for Asset Managers, Dr de Prado (Elements in Quantitative Finance)

- Financial Event Prediction using Machine Learning, Dr Snow (SSRN)
- DeltaPy: A Framework for Tabular Data Augmentation in Python, Dr Snow (SSRN)
- Machine Learning in Asset Management—Part 1, Dr Snow (Trading Strategies)
- Machine Learning in Asset Management—Part 2, Dr Snow (Weight Optimisation)
- Advances in Financial Machine Learning, Dr de Prado (Wiley)
- History of Machine Learning in Finance and Economics, Dr Snow (article)
- The Front Office: A Hedge Fund Guide for Retail, Day Traders, and Aspiring Quants

Letter Grades

Letter grades for the entire course will be assigned as follows. I **only** work with **your percentages score in assignments** (your score (1-4) will only show at the end).

Letter Grade	# Points	Aa Percent
A	4	95% and higher
A-	3.67	90.0 – 94.99%
B+	3.33	87.5% - 89.99%
B	3	82.5% - 87.49%
B-	2.67	80% - 82.49%
C+	2.33	77.5% - 79.99%
C	2	72.5% - 77.49%
C-	1.67	70% - 72.49%
D+	1.33	67.5% - 69.99%
D	1	62.5% - 67.49
D-	0.67	60% - 62.49%
F	0	59.99% and lower

Course Materials

Textbooks & Materials

- Course website: ml-quant.com/class-2023 (notion access provided in class).

Resources

- **Access your course materials**: NYU Classes (nyu.edu/its/classes)
- **Databases, journal articles, and more**: Bern Dibner Library (library.nyu.edu) ● **Obtain 24/7 technology assistance:** Tandon IT Help Desk (, 646.997.3123) NYU IT Service Desk (, 212-998-3333).

  soehelpdesk@nyu.edu

  AskIT@nyu.edu
- Use **Google Scholar** https://scholar.google.com/ for academic papers, patents and legal cases.
- For your reports and presentations, learn how to use the bibliographic management tools: see https://guides.nyu.edu/citations/tools . I use Mendeley to organize citations, format bibliographies & citations, and more. See https://guides.nyu.edu/mendeley.

Academic Misconduct

You are not allowed to present other people's work as your own. Summarize in your own words ("paraphrase"), quote, cite, and provide a professionally formatted reference.

Copying violates professional standards. Review the NYU and Tandon Codes of Conduct at

- https://engineering.nyu.edu/sites/default/files/2018-06/code-conduct2-2-16.pdf
- http://www.nyu.edu/about/policies-guidelines-compliance/policies-andguidelines/educational-and-research-uses-of-copyrighted-materials-policy-st.html For more information: Consult point 5 on http://inductive.net/fe/news/FAQ.htm .

1. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.
2. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:
    a. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
    b. Fabrication: including but not limited to, falsifying experimental data and/or citations.
    c. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
    d. Unauthorized collaboration: working together on work that was meant to be done individually.
    e. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
    f. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

Disability Disclosure Statement

Academic accommodations are available for students with disabilities. Please contact the **Moses Center for Students with Disabilities** (212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center as early as possible in the semester for assistance. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 2nd floor.

Syllabus Inclusion Statement

The NYU Tandon School values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. If this standard is not being upheld, please feel free to speak with me.