

MACHINE LEARNING IN QUANTITATIVE FINANCE

Academic Year: 2026/2027**Semester:** S1**Number of Students:** 5

Seminar sessions will be held remotely. Mandatory participation in all sessions.

INSTRUCTOR

Dan H. Tran

CONTACTS

Email: dtran@ucp.pt

Office: 5316

Website: dan-tran.eu

BIOGRAPHY

Dan Tran is an invited assistant professor at the Católica Lisbon School of Business & Economics. He holds a Ph.D. in financial economics, and a M.Sc. in Risk Engineering. His research focuses on computational modelling and machine learning in Finance. His works cover topics such as financial stability, network externalities, and synthetic data. He has experience working with ML systems in industry applications.

SEMINAR OVERVIEW

Machine Learning (ML) is a powerful toolbox that combines data science and computational methods. ML techniques are particularly useful in finance where practitioners consistently work with high-dimensional data and non-linear relationships. However, financial data exhibit low signal-to-noise ratio and strong inter-dependencies. Therefore, financial ML applications are not straightforward and challenging.

This seminar will help students to gain deep insights into deploying ML for Finance. Research topics will focus on 2 themes: (1) Natural Language Processing (NLP) with text data, and (2) asset pricing with a focus on factor models and factors timing. Candidates can make proposal for new topics, to be validated by the supervisor.

Placement of past successful candidates: asset management (Citi, DWS, Pictet, UBS), data science (Mastercard, Norway Statistics), and M&A, consulting (e.g. Triton, Deloitte, BNP).

One important ingredient for successful projects is the experimental mindset: evidence-based thinking and fast iteration cycles.

SEMINAR OBJECTIVES

The goal is to deliver *end-to-end solutions to relevant issues in finance, with explicit economic and research value*.

On the technical side, participants will become operational in ML with regards to: process data, build and train ML models, evaluate how these models contribute to economic outcomes or answer relevant questions.

On the analytical side, participants will develop critical thinking, improve the capacity to solve challenging problems, communicate complex ideas in a comprehensible manner, efficient project management cycle.

Past successful projects have demonstrated the following characteristics: technically sound, innovative problem solving, clear communication, good time management.

SEMINAR CONTENT

This seminar consists of group sessions and individual sessions. All materials and interactions are in English.

Group sessions include: setup workflow and research topics, review past projects, internal ML challenge.

On the individual basis, participants will review relevant literature, acquire necessary technical knowledge in both finance and ML, collect and analyze data, evaluate models and results. Individual support is available on demand at this stage. Total individual meeting will not exceed 4 hours. Continuous support is provided through Notion and Discord servers.

Participants are expected to deliver specific output at scheduled milestones, make presentations, and give feedback on past and current works.

In Finance, ML is currently used in asset pricing and NLP with text data. The following articles can be used for inspiration:

- Gu, Kelly & Xiu (2020). "Empirical asset pricing via machine learning". The Review of Financial Studies, 33(5), 2223-2273.
- Li, K., Mai, F., Shen, R., & Yan, X. (2021). "Measuring Corporate Culture Using Machine Learning". The Review of Financial Studies, 34(7), 3265-3315.
- Kaniel, R., Lin, Z., Pelger, M., & Van Nieuwerburgh, S. (2023). "Machine-learning the skill of mutual fund managers". Journal of Financial Economics, 150(1), 94-138.
- Bryan Kelly and Dacheng Xiu (2023), "Financial Machine Learning", NBER (A survey).

For NLP, we use corporate filings (10K, 10Q, 8K) and earning calls transcripts. For financial markets, we use WRDS and available datasets from well-known finance papers. Other topic could be viable if the candidate demonstrates solid domain knowledge.

REQUIRED BACKGROUND

Basic knowledge of ML and Python are required. Good command of statistics, programming and experience with handling data are recommended. As a checkpoint, students are required to complete an exercise to validate their skills.

Strong interests in financial markets, prior knowledge and domain expertise in any area of finance may constitute an advantage.

Students are expected to work full-time on the project for 4 months. One extension period is supported, but no further.

SCHEDULE

Provisional schedule:

- August: invite to Discord server, reading materials.
- September 1st-18th: 3x group sessions, Friday 2-4pm, mandatory presence.
- September 25-27th: Kaggle/internal ML challenge
- October-December: individual sessions, on-demand
- Total time: 20 weeks

Milestones:

1. Research proposal (20/09)
2. Research plan – Data validation (04/10)
3. Proof-of-concept (25/10)
4. First draft (29/11)
5. Final draft (28/12)
6. Final delivery & defense authorization (04/01)

All deadlines and milestones are mandatory. Failure to respect these milestones could result in termination of the project.

DEADLINES

The full list of dates and deadlines will be available [here](#).

DISSERTATION GUIDELINES

The students can also find information regarding the dissertation content and standards, format, length, and submission on the Dissertation Guidelines [here](#).

BIBLIOGRAPHY

Scientific articles in peer-reviewed, including but are not limited to: Journal of Finance, Journal of Financial Economics, Review of Financial Studies, Journal of Financial and Quantitative Analysis, Review of Finance, Journal of the American Statistical Association, Journal of Econometrics, Econometrica, Journal of Applied Econometrics, Review of Economics and Statistics, American Economic Review, Journal of Monetary Economics, Review of Economic Studies, Journal of Business and Economic Statistics, Management Science; Journal of Machine Learning Research, Science, Nature. Proceedings of ML conferences: NeurIPS, ICML, ICLR, AAAI, KDD, EMNLP.

Additional reading:

Murphy, K. (2013). “Machine learning: a probabilistic approach”. MIT Press – selected chapter.

Burkov, A. (2020). “The hundred-page machine learning book”.

ADDITIONAL RESOURCES

Cloud-computing resources and datasets could be provided upon request, depending on the project.

CODE OF CONDUCT AND ETHICS

Católica Lisbon School of Business and Economics is a community of individuals with diverse backgrounds and interests who share certain fundamental goals. A crucial element to achieve these goals is the creation and maintenance of an atmosphere contributing to learning and personal growth for everyone in the community. The success of CATÓLICA-LISBON in attaining its goals and in maintaining its reputation of academic excellence depends on the willingness of its members, both collectively and individually, to meet their responsibilities.

Along with all the other members of our community, students are expected to follow professional standards and CATÓLICA-LISBON standards of Academic Integrity. Some details should be mentioned here: Please arrive on time for class with uninterrupted attendance for the duration of the class. Signing attendance sheet for anyone else in the class constitutes fraud and a violation of the CLSBE code of conduct. Use of computers and other electronic devices during the class is not allowed, unless expressly

requested by the instructor of the course. Students who persistently act in a disruptive and disrespectful manner during the class session may be invited to leave.

Students are expected to behave at all times according to the fundamental principles of academic integrity, including honesty, trust, fairness, respect, and responsibility. In particular,

- a. In **individual graded assignments** of any type, students may not collaborate with others or use any materials without explicit permission from the instructor of the course;
- b. In **group assignments** and reports, all students listed as authors should have performed a substantial amount of work for that assignment;
- c. It is dishonest to fabricate or falsify data in experiments, surveys, papers, reports or other circumstances; fabricate source material in a bibliography or “works cited” list; or provide false information in other documents in connection with academic efforts;
- d. **Plagiarizing**, i.e. “to steal and pass off the ideas or words of another as one’s own and or to use another’s production without crediting the source” (Merrian-Webster Dictionary) is an Academic Integrity breach. It can be avoided by using proper methods of documentation and acknowledgement. Visit this guide for additional resources on how to avoid plagiarism in your written submissions <http://en.writecheck.com/plagiarism-guide>
- e. In **exams** students must not receive or provide any unauthorized assistance. During an examination, students may use only material and items authorized by the faculty. Use of smartwatches or other communication devices is not permitted during the exam.

Academic integrity breaches will be dealt with in accordance with the school’s code of Academic Integrity: <https://www.clsbe.lisboa.ucp.pt/system/files/assets/files/academicintegritycode.pdf>