Discipline:

Methods & Finance

1. Language

English

2. Title

Machine Learning in Finance

3. Lecturer

Prof. Dr. Martin Hibbeln, Mercator School of Management, Universität Duisburg-Essen
(martin.hibbeln@uni-due.de)

4. Date and Location

21.04.2026 – 02.06.2026

The course will be offered online via Zoom. After the kickoff session in week 1, weeks 2 – 5 each include a 4-hour lecture and a 4-hour computer lab session. Afterwards, there will be a small Kaggle coding competition with group presentations of the results.

5. Course Description

5.1 Abstract and Learning Objectives

Recent developments in analytical techniques, massive increases in computing power, and growing amounts of data have led to an unprecedented rise in machine learning applications across many areas of finance, such as asset pricing (Gu et al. 2020; Bianchi et al. 2021), earnings forecasts (van Binsbergen et al. 2022), and credit scoring (Fuster et al. 2021). Clearly, machine learning has become an important skill for any aspiring financial researcher.

The goal of this course is to provide participants with a solid methodological foundation in machine learning models, demonstrate how these methods can be applied to address financial research questions, and enable the participants to analyze both structured and unstructured data using machine learning in R.

At the end of the course, participants will apply their knowledge in a Kaggle coding competition, where they team up with fellow participants to develop the most effective prediction model for a prediction task in a financial context.

5.2 Content

1. Machine Learning Fundamentals: Prediction Models and Model Evaluation
2. Subset Selection, Shrinkage, and Dimension Reduction
3. Moving Beyond Linearity
4. Unstructured Data

5.3 Schedule (including start and end time)

Before each session, participants are required to watch the accompanying videos from James et al. (2023), which we will provide. These videos give participants a methodological understanding of how the machine learning models work. This allows us to focus on their application in a financial context during the lecture and their implementation in R in the computer lab.

Topic 1: Machine Learning Fundamentals: Prediction Models and Model Evaluation

Topic 2: Subset Selection, Shrinkage, and Dimension Reduction

Topic 3: Moving Beyond Linearity

Topic 4: Unstructured Data

Date	Time	Session
21.04.2026	14:00 – 15:30	Kickoff Session
28.04.2026	14:00 – 18:00	Topic 1: Lecture
29.04.2026	09:00 – 12:00	Topic 1: Lab Session
05.05.2026	13:00 – 16:00	Topic 2: Lecture
06.05.2026	09:00 – 12:00	Topic 2: Lab Session
12.05.2026	14:00 – 17:30	Topic 3: Lecture
13.05.2026	09:00 – 12:00	Topic 3: Lab Session
19.05.2026	13:00 – 16:00	Topic 4: Lecture
20.05.2026	09:00 – 12:00	Topic 4: Lab Session + Introduction Kaggle Competition
02.06.2026	10:00 – 16:00	Presentations Coding Challenge

5.4 Course format

Lectures and computer lab sessions

6. Preparation and Literature

6.1 Prerequisites

- Foundations in statistics (descriptive statistics, linear regression)
- Basic knowledge of programming in R

6.2 Essential Reading Material

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). An Introduction to Statistical Learning. 2nd Edition. New York: Springer.

6.3 Additional Reading Material

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. 2nd Edition. New York: Springer.

Topic 1: ML Fundamentals: Prediction Models and Model Evaluation

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33, 2223–2273.

Gürtler, M., Hibbeln, M. T., & Usselmann, P. (2018). Exposure at Default Modeling–A Theoretical and Empirical Assessment of Estimation Approaches and Parameter Choice. Journal of Banking & Finance, 91, 176–188.

Topic 2: Subset Selection, Shrinkage and Dimension Reduction

Croux, C., Jagtiani, J., Korivi, T., & Vulanovic, M. (2020). Important Factors Determining Fintech Loan Default: Evidence from a Lendingclub Consumer Platform. Journal of Economic Behavior & Organization, 173, 270–296.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33, 2223–2273.

Topic 3: Moving Beyond Linearity

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33, 2223–2273.

Topic 4: Unstructured Data

Lang, M., & Stice-Lawrence, L. (2015). Textual Analysis and International Financial Reporting: Large Sample Evidence. Journal of Accounting and Economics, 60, 110–135.

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager Sentiment and Stock Returns. Journal of Financial Economics, 132, 126–149.

Obaid, K., & Pukthuanthong, K. (2022). A Picture Is Worth a Thousand Words: Measuring Investor Sentiment by Combining Machine Learning and Photos From News. Journal of Financial Economics, 144, 273–297.

6.4 To prepare

Participants are expected to watch selected videos accompanying James et al. (2023), which will be provided by us, before each session. Moreover, we expect the participants to have knowledge of basic statistical concepts and have some experience with the programming language R.

7. Administration

7.1 Max. number of participants

20 participants

7.2 Assignments

None

7.3 Exam

After the course, we will host a coding competition via Kaggle where participants will team up in groups and develop predictive models for a prediction task in a financial context. The participants submit and present their results via Zoom around 2 weeks after the end of the course period.

7.4 Credits

The course corresponds to a scope of 6 ECTS.

8. Working Hours

	Hours
Preparation of programming in R	30h
Course preparation / reading material	20h
Watching videos in preparation for each session	10h
Active participation in class during lectures and lab sessions	40h
Follow-up work (read up on topics of the day to deepen understanding)	40h
Kaggle coding competition and presentation	40h
Sum	**180 h**