UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Chair of Finance

Prof. Dr. Martin Thomas Hibbeln

Room LH 005
Phone 0203 / 379 – 1784
E-Mail finance@uni-due.de
Website udue.de/finance

Lecture Summer Semester 2025

Machine Learning in Finance

In-person Meetings & Videos:
Monday & Thursday, 2–6 p.m. (c.t.): LB 335 (Monday) & LC 134 (Thursday)
Kickoff Session: 14.04.2025

Date	Topic
14.04.2025	Kickoff Session
24.04.2025 (Lecture) 28.04.2025 (Lab)	Machine Learning Fundamentals: Prediction Models and Model Evaluation
05.05.2025 (Lecture) 08.05.2025 (Lab)	Subset Selection, Shrinkage, and Dimension Reduction
12.05.2025 (Lecture) 15.05.2025 (Lab)	Moving Beyond Linearity
19.05.2025 (Lecture) 22.05.2025 (Lab)	Unstructured Data
12.06.2025	Kickoff Coding Challenge
07.07.2025	Deadline Term Paper
14.07.2025 & 17.07.2025	Presentations

On the following pages you will find a detailed description of the content of each session and references. The latter can be downloaded from Moodle. Please also note the following organizational notes.

Schedule In-person Meetings:

	Monday (LB 335)	**Thursday** (LC 134)
Week 1: 14.04. – 18.04. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Kickoff Session	—
Week 2: 21.04. – 25.04. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	(public holiday)	Lecture: Topic 1
Week 3: 28.04. – 02.05. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Lab Session: Topic 1	(public holiday)
Week 4: 05.05. – 09.05. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Lecture: Topic 2	Lab Session: Topic 2
Week 5: 12.05. – 16.05. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Lecture: Topic 3	Lab Session: Topic 3
Week 6: 19.05. – 23.05. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Lecture: Topic 4	Lab Session: Topic 4
Week 9: 09.06. – 13.06. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.		Kickoff Coding Challenge
Week 14: 14.07. – 18.07. 2 p.m. – 4 p.m. 4 p.m.– 6 p.m.	Presentations	Presentations

Organizational Notes

The course *Machine Learning in Finance* integrates home study with videos, lectures, and computer lab sessions. Before each lecture, students are required to watch instructional videos that cover the methodological foundations, which are then applied in class. During the lectures, we discuss recent empirical research that uses machine learning techniques to address finance-related problems. In the computer lab sessions, students gain hands-on experience by coding state-of-the-art machine learning models and applying them to empirical data.

The course is designed for Master students in Finance and Data Analytics and Management and Economics (UDE), Master students enrolled in Econometrics (UAR) and students from RGS Econ.

The assessment is based on a term paper and a corresponding presentation in which students apply their machine learning knowledge to a financial prediction task. The goal is to implement multiple predictive models to achieve a high predictive accuracy. In addition to the coding component, students must document their approach in a short paper and present their results to the chair on a scheduled date. The coding challenge will be posted within a separate kickoff session at 12.06. 2 p.m. Students will work on the coding challenge and the term paper in groups. The term paper is expected to be submitted by 07.07., 12 p.m. The presentation will take place on 14.07.2025 & 17.07.2025 from 2 p.m. – 6 p.m. Attendance is required for all presentations. The second day of presentations will be canceled in case all presentations can be accommodated on the first day.

Note that the videos are based on the following book, which accompanies the entire lecture:
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. 2nd Edition. New York: Springer.

In the following, we provide a brief overview of the topics covered in the lecture as well as the empirical papers discussed in each of the topics.

Brief description of the topics

Topic 1: ML Fundamentals: Prediction Models and Model Evaluation

Students will gain an understanding of the fundamental elements of Machine Learning, including simple linear models for regression and classification. They will explore how prediction tasks differ from statistical inference. A major focus will be on model evaluation, which will be explored from several angles, including resampling techniques, OOS vs. OOT distinctions, and evaluation metrics.

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.

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

Topic 2: Subset Selection, Shrinkage and Dimension Reduction:

Students will be introduced to various types of linear models designed to mitigate overfitting in training data. While subset selection methods such as Forward, Backward, and Stepwise Selection will be briefly covered, the main focus will be on state-of-the-art regularization techniques such as LASSO and Ridge, and dimension reduction methods such as PCA and PLS. Suitable applications for these techniques wills also be discussed.

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

We will extend beyond linear models that enforce a strictly linear relationship between input and output. Students will explore a variety of nonlinear methods that define the current state-of-the-art in machine learning. The course will cover tree-based approaches, including Decision Trees, Bagging, Boosting, and Random Forests, as well as Neural Networks. In addition, students will be introduced to feature importance techniques, which have gained prominence in addressing the black-box problem of ML models.

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

Topic 4: Unstructured Data

In the final session, students will be introduced to unstructured data, which makes up more than 80% of the data available. They will learn methods for analyzing both text and image data. For text analysis, topics will include sentiment analysis and readability assessment. For image analysis, students will build on their prior knowledge of neural networks by exploring convolutional neural networks, the state-of-the-art approach to image processing and analysis.

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.