Ratmir Miftachov

Citizenship: German | contact@miftachov.com | GitHub | Google Scholar | LinkedIn

Education

Humboldt University of Berlin — Berlin, Germany
PhD Candidate in Mathematical Statistics — *May 2021 - expected: April. 2026*
- Supervisor: Wolfgang Härdle

Princeton University — Princeton, United States
Visiting Researcher, Operation Research and Financial Engineering Department — *May 2023, Aug. - Dec. 2024*
- Invited by Jianqing Fan

University of Mannheim — Mannheim, Germany
MSc in Economics with Distinction (Major: Statistics) — *Sept. 2018 - Sept. 2020*
- Semester abroad at ETH/UZH (Zurich, Switzerland).

University of Cologne — Cologne, Germany
BSc with Distinction (Major: Statistics) — *Sept. 2015 - Febr. 2018*
- Finished one semester faster

Work experience

Applied Scientist Intern, Amazon — Paris, France
Research paper on deep learning for tabular data — *July 2025 - Dec. 2025*

Research Associate, Department of Mathematics
SFB 1294 'Nonlinear statistical inverse problems', HU Berlin — *Aug. 2023 - July 2025*

Supervision Master's thesis
Mustafa Suman on 'Early Stopping for Random Forest Classifier' (together w/ Markus Reiß) — *2024*

Teaching Assistant
'Statistical Learning and Data Science' (Master's level), HU Berlin — *Spring, Fall 2022*

Associated researcher, DeSBi
'Deep Learning and Statistics Towards Understanding Structured Biomedical Data', HU Berlin — *Since Mar. 2023*

Research Associate, Department of Economics
IRTG 1792 'High-dimensional nonstationary time series', HU Berlin — *May 2021 - Feb. 2023*

Research Intern, German Federal Bank — *Summer 2018*
Department General Statistics, Frankfurt am Main, Germany — *[slides][code]*

Research Projects

Interpretability of foundation models for tabular data
To be submitted
- Research project at Amazon. Case-based reasoning for large-scale transformers for tabular data.

Early Stopping for Regression Trees
Under revision — *[paper][slides][code]*
- Proved theoretical guarantees for a novel data-driven early stopping rule for the regression tree. Demonstrated its ability to significantly reduce computational costs compared to state-of-the-art methods.

EarlyStopping: Implicit Regularization for Iterative Learning Procedures in Python
Revise and resubmit at Journal of Statistical Software — *[paper][code]*
- Implemented a Python library for early stopping techniques for different algorithms, including a gradient descent variant, conjugate gradient, tSVD, L2-boosting, and the regression tree.

Early Stopping for Random Forest Classifier
Work-in-Progress — *[code]*
- Extended my previous research to the random forest classifier. The early stopped forest is computationally efficient, has significantly fewer nodes, and its prediction performance is on par with the deeply grown forest.

Shapley Curves: A Smoothing Perspective
Published at Journal of Business and Economic Statistics *[paper][slides][code]*

- Derived minimax rates for nonparametric Shapley curves as a variable importance measure and established a novel wild bootstrap procedure for finite sample inference.

Risk-Premia in the Bitcoin Market
Revise and resubmit at Journal of Business and Economic Statistics *[paper][slides][code]*

- Analyzed Bitcoin option data through the nonparametric pricing kernel and identified a different risk appetite compared to the S&P 500 based on a novel clustering algorithm.

Awards & Miscellaneous

PhD Scholarship of the German Academic Scholarship Foundation (Studienstiftung)
Merit-based Scholarship awarded to the top 0.5% of German students *Since 2023*

- 4 years; approx. EUR 100,000

Dean's award for outstanding academic achievement
Awarded to the top 5% of 479 students in the Bachelor's program *2017*

Talks

Mathematical Statistics Seminar (invited) Berlin, Germany
Weierstrass Institute of Applied Analysis and Stochastics *May. 2025*

Workshop on Early Stopping (invited) Paris, France
Institut Henri Poincaré *Mar. 2025*

Statistics Lab Seminar (invited) Princeton, United States
Princeton University *Oct. 2024*

International Conference on Computational Statistics (invited) London, United Kingdom
University of London *Aug. 2023*

Conference Statistical Foundations of Data Science Princeton, United States
Princeton University *May 2023*

Conference Recent Advances in Statistics and Data Science Rutgers NJ, United States
Rutgers University *May 2023*

Conference Statistics of Machine Learning (invited) Prague, Czech Republic
Charles University *Oct. 2022*

Frequently Used Skills

Languages: German (native), English, Russian
Software: Python (6 research projects, 1 statistical package), R (2 research projects, pre-PhD studies), Matlab (pre-PhD studies), Git (used across all projects), Cursor/Roo Code
Mathematical statistics: likelihood inference, parametric and nonparametric regression (e.g. splines, kernel regression), experimental design, statistical hypothesis testing, bootstrap inference (finite and asymptotic), minimax theory, probability theory, time series estimation, GMM, Bayesian statistics (MCMC, variational inference, PPDs)
Statistical learning: tree based algorithms (CART, bagged trees, random forest), boosting, classification (e.g. logistic regression, k-NN, SVM), causal inference (Rubin's causal model, Pearl's structural causal model), cluster analysis, GANs, regularization techniques (e.g. LASSO, ridge, early stopping), dimensionality reduction (PCA, UMAP, LLE, t-SNE), LSTM, explainable ML (e.g. variable importance, SHAP), experience with transformers and LLMs, Foundational models for tabular data (TabPFN, TabICL, etc.), deep learning (PyTorch)