

Nikita Kozodoi

ML researcher & PhD candidate working on the frontier of research and business. With certified AWS ML knowledge, practical experience in different ML areas and strong communication skills, I am passionate about using AI to improve decision-making. My recent projects focus on fair ML for credit scoring and DL for image and text classification.

 kozodoi.me (portfolio & ML blog)

 linkedin.com/in/kozodoi

 github.com/kozodoi

 n.kozodoi@icloud.com

Experience

- 04/2018 – present **Data Science Research Associate**, *Monedo / Humboldt University*, Berlin.
- developed sample bias mitigation algorithms to improve model profitability by up to 3%
 - benchmarked fair ML algorithms to identify methods enhancing fairness at the smallest cost
 - reduced data acquisition costs by up to 25% using multi-objective feature selection methods
 - published 5 ML papers with over 65 total citations and delivered 5 conference talks [\[Papers\]](#)
- 09/2018 – 08/2020 **Project-Based Data Scientist**, *Simply Rational*, Berlin, freelance.
- identified customer segments and prototyped segment prediction models from incomplete data
 - determined suitable treatments from A/B tests results and presented findings to corporate clients
- 07/2016 – 03/2018 **Research Assistant**, *Max Planck Institute for Human Development*, Berlin.
- benchmarked ML algorithms against simple heuristics and analyzed key performance drivers

Education

- 04/2018 – 03/2022 **Ph.D. in Machine Learning, Information Systems**, *Humboldt University*, Berlin.
Research on ML for credit scoring. Teaching ML/DL. Supervising student theses and projects.
- 10/2015 – 12/2017 **M.Sc. in Economics, Quantitative Methods**, *Humboldt University*, Berlin, GPA: 1.30 (A).
Focused on Data Science and Predictive Analytics. Thesis topic: ML for box office forecasting.
- 09/2010 – 06/2014 **B.Sc. in Economics**, *Higher School of Economics*, Moscow, GPA: 1.31 (A).
Focused on Statistics and Econometrics. Top-1 in the 4-year student grade rating. Diploma with Honors.

Certifications

- 01/2022 **AWS Certified Machine Learning**, *AWS*. Training, tuning and deploying ML models on AWS. [\[Certificate\]](#)
- 06/2020 **Deep Learning Nanodegree**, *Udacity*. Implemented CNNs, RNNs and LSTMs in PyTorch. [\[Certificate\]](#)
- 05/2020 **Machine Learning Engineer Nanodegree**, *Udacity*. Deployed ML models on AWS SageMaker. [\[Certificate\]](#)





Selected ML Projects





- 05/2021 – 08/2021 **Text Readability Prediction**. Developed a PyTorch pipeline for text complexity prediction with transformers. Deployed a web app for scoring custom texts. Placed top-9% in the Kaggle competition. [\[App\]](#) [\[GitHub\]](#)
- 03/2021 – 06/2021 **Image-to-Text Molecular Translation**. Translated molecule images into chemical formulas with PyTorch. Built a CNN-LSTM architecture for image captioning. Placed top-5% in the Kaggle competition. [\[GitHub\]](#)
- 09/2019 – 04/2021 **Fairness: Package for Fair ML**. Building and maintaining an R package with over 16k total downloads. Implemented calculation and visualization of fairness metrics. Wrote documentation. [\[Blog post\]](#) [\[GitHub\]](#)
- 09/2018 – present **Kaggle Master**. Won 18 competition medals. Top-1% in Competitions, Notebooks and Discussion. [\[Projects\]](#)

Selected ML Publications

- 06/2020 – present **ML Blog**. Publishing ML/DL tutorials and project overviews. Reached over 7k monthly page views. [\[Blog\]](#)
- 11/2021 **Fairness in credit scoring: Assessment, implementation and profit implications**, *EJOR*. [\[Paper\]](#) [\[GitHub\]](#)
- 04/2020 **Shallow self-learning for reject inference in credit scoring**, *ECML PKDD*. [\[Paper\]](#) [\[Meetup talk\]](#)
- 05/2019 **A multi-objective approach for profit-driven feature selection in credit scoring**, *DSS*. [\[Paper\]](#)

Technical Stack

Tools  Python, R, SQL (*interm*)  VSCode, Jupyter  AWS Cloud (SageMaker, EC2)  GitHub, Trello

Packages  Pandas, NumPy  PyTorch, TensorFlow (*interm*)  Scikit-learn, XGBoost  Timm, Huggingface

Soft Skills

- Languages **English** (C2), **German** (C1), **Russian** (native)
- Key skills
- Strong motivation to learn and improve
 - Excellent communication skills [\[Public talks\]](#)
 - Attention to detail and scientific mindset
 - Rich experience of working in team projects