

# Nikita Kozodoi

ML/DL researcher working on the frontier between research and business. With certified ML expertise, rich experience in competitive ML, and strong coding and public talking skills, I am enthusiastic about using AI to improve decision-making. My latest projects focus on ML for tabular and computer vision problems.

 <https://kozodoi.me> (portfolio & ML blog)

 [linkedin.com/in/kozodoi](https://linkedin.com/in/kozodoi)

 [github.com/kozodoi](https://github.com/kozodoi)

 [n.kozodoi@icloud.com](mailto:n.kozodoi@icloud.com)

## Experience

- 04/2018 – present **Data Science Research Associate**, *Monedo / Humboldt University*, Berlin.
- developed sample bias correction algorithms to improve model profitability by up to 3%
  - reduced data acquisition costs by up to 25% using multi-objective feature selection
  - published 5 research papers with 65 citations and delivered 5 conference talks [\[Papers\]](#)
- 09/2018 – 08/2020 **Project-Based Data Scientist**, *Simply Rational*, Berlin.
- identified customer segments for targeting and predicted clusters from incomplete data
  - determined suitable treatments from A/B tests and presented findings to corporate clients
- 07/2016 – 03/2018 **Research Assistant**, *Max Planck Institute for Human Development*, Berlin.
- benchmarked ML algorithms against managerial heuristics and analyzed key performance drivers

## Education

- 04/2018 – present **PhD in Machine Learning, Information Systems**, *Humboldt University*, Berlin.  
Research on ML in credit scoring. Teaching Data Science and supervising student theses.
- 10/2015 – 12/2017 **M.Sc. in Economics, Quantitative Methods**, *Humboldt University*, Berlin, GPA: 1.30 (A).  
Focused on ML and Data Science. Double-degree program with HSE Moscow.
- 09/2010 – 06/2014 **B.Sc. in Economics**, *Higher School of Economics*, St. Petersburg, GPA: 1.31 (A).  
Focused on Statistics and Econometrics. Top-1 in the 4-year student rating.

## Certifications

- 01/2022 **AWS Certified Machine Learning**, *Amazon*. Training, tuning and deploying ML models on AWS. [\[Certificate\]](#)
- 06/2020 **Deep Learning Nanodegree**, *Udacity*. Implemented CNNs, RNNs and GANs in PyTorch. [\[Certificate\]](#)
- 05/2020 **Machine Learning Engineer Nanodegree**, *Udacity*. Deployed ML models on AWS SageMaker. [\[Certificate\]](#)
- Courses with certificates: Data Structures, Algorithmic Toolbox, SQL for Data Science.











## Selected ML Projects

- 06/2020 – present **Machine Learning Blog**. Implementing and hosting ML blog with over 7k monthly page views.  
Publishing ML/DL tutorials, Kaggle competition solutions and project overviews. [\[Blog\]](#) [\[GitHub\]](#)
- 05/2021 – 08/2021 **Text Readability Prediction**. Developed a PyTorch pipeline for text complexity prediction with transformers.  
Built an interactive web app for text scoring. Reached top-9% in the Kaggle competition. [\[Web app\]](#) [\[GitHub\]](#)
- 03/2021 – 06/2021 **Image-to-Text Molecular Translation**. Translated molecule images into chemical formulas with PyTorch.  
Implemented a CNN-LSTM architecture for image captioning. Reached top-5% in the Kaggle competition. [\[GitHub\]](#)  
Kaggle Master. Won 18 competition medals. Top-1% in Competitions, Notebooks and Discussion. [\[Profile\]](#) [\[Portfolio\]](#)

## Selected ML Publications

- 06/2021 **Fairness in credit scoring: Assessment, implementation and profit implications**, *European Journal of OR*.  
Benchmarked fair ML algorithms and analyzed the profit-fairness trade-off. [\[Paper\]](#) [\[GitHub\]](#)
- 04/2020 **Shallow self-learning for reject inference in credit scoring**, *European Conference on ML and PKDD*.  
Proposed a self-training based approach for bias correction in credit scoring. [\[Paper\]](#) [\[Slides\]](#)
- 09/2019 **A multi-objective approach for profit-driven feature selection in credit scoring**, *Decision Support Systems*.  
Developed a profit-driven feature selection framework using multi-objective genetic algorithms. [\[Paper\]](#)

## Technical Stack

Tools	 Python, R, SQL ( <i>interm</i> )	 VSCode, Jupyter	 AWS Cloud (SageMaker, EC2, Lambda)	 Git, Trello
Packages	 Pandas, NumPy	 Scikit-learn, XGBoost, LightGBM, Optuna	 Matplotlib, Seaborn	
	 PyTorch, TensorFlow ( <i>interm</i> )	 Timm, Albumentations	 Huggingface	

## Soft Skills

- Languages **English (C2), German (C1), Russian (native)**
- Key skills
- Rich experience of working in team projects
  - Excellent presentation skills [\[Public talks\]](#)
  - Scientific mindset and strong motivation to learn
  - Business-oriented thinking