Nikita Kozodoi

ML/DL researcher with expertise in competitive ML and economic background

- n.kozodoi@icloud.com
- https://linkedin.com/in/kozodoi
- https://kozodoi.me (project portfolio & ML blog)
- https://github.com/kozodoi

Experience

04/2018 – present **Data Science Research Associate**, *Monedo / Humboldt University*, Berlin.
- developing solutions to improve ML model training and evaluation under sampling bias
- implementing fair ML algorithms and investigating profit-fairness trade-off in credit scoring models
- optimizing data acquisition costs with multi-objective profit-driven feature selection

09/2018 – 08/2020 **Project-Based Data Scientist**, *Simply Rational*, Berlin.
- worked on multiple projects, including customer segmentation and treatment effect estimation
- preprocessed data and developed project-specific ML pipelines and prototypes for corporate clients

07/2016 – 03/2018 **Research Assistant**, *Max Planck Institute for Human Development*, Berlin.
- benchmarked managerial heuristics against ML algorithms in customer classification problems
- analyzed factors affecting the predictive performance on marketing data sets

Education

04/2018 – present **PhD in Information Systems, Machine Learning**, *Humboldt University*, Berlin.
Research on ML/DL applications for credit scoring. Teaching ML and supervising M.Sc. dissertations.

10/2015 – 12/2017 **M.Sc. in Economics and Management Science**, *Humboldt University*, Berlin, GPA: 1.30 (A).
Focused on Machine Learning 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, Time Series and Econometrics. Top-1 in the 4-year student grade rating.

Certificates

11/2021 **SQL for Data Science**, *UC Davis*. Learned SQL for data extraction and manipulation

06/2020 **Deep Learning Nanodegree**, *Udacity*. Implemented CNNs, RNNs and GANs in PyTorch [Certificate]

05/2020 **Machine Learning Engineer Nanodegree**, *Udacity*. Deployed ML/DL models on AWS Sagemeaker [Certificate]

Technical Stack

Tools	Python, R, SQL *(interm)*	VSCode, Jupyter	AWS (Sagemaker, EC2), Neptune.ai	Git, Trello
Packages	Pandas, NumPy	Scikit-learn, XGBoost, LightGBM, Hyperopt	Matplotlib, Seaborn
	PyTorch, TensorFlow *(interm)*	Timm, Albumentations	HuggingFace
Applications	classification, regression, clustering, dimensionality reduction	sentiment analysis, text classification	
	image classification, object detection, image captioning		

Selected ML Competitions

08/2021 **CommonLit Readability Prize**, top-9%. Predicted text readability with transformers [GitHub] [Web app]

06/2021 **BMS Molecular Translation**, top-5%. Built CNN-LSTM architecture for image-to-text translation [GitHub]

02/2021 **Cassava Leaf Disease Classification**, top-1%. Identified plant diseases with deep learning [GitHub]

Kaggle Master. Won 17 competition medals. Top-1% in Competitions, Notebooks and Discussion [Profile]

Selected ML Publications

06/2021 **Fairness in credit scoring: Assessment, implementation and profit implications**
European Journal of Operational Research [doi.org/10.1016/j.ejor.2021.06.023] [GitHub]

04/2020 **Shallow self-learning for reject inference in credit scoring**
European Conference on ML and PKDD [doi.org/10.1007/978-3-030-46133-1] [Slides]

09/2019 **A multi-objective approach for profit-driven feature selection in credit scoring**
Decision Support Systems [doi.org/10.1016/j.dss.2019.03.011]

Further Skills

Languages **English** (C2), **German** (C1), **Russian** (native speaker)

Key skills
- Passionate about machine learning and data
- Inspired by using ML to improve decision-making
- Excellent presentation skills [My talks]
- Scientific mindset and strong motivation to learn

Berlin, 08.11.2021