# Nikita Kozodoi

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

## Experience

**04/2018 – present** — **Data Science Research Associate**, *Monedo / Humboldt University*, Berlin.
- developing solutions to mitigate the impact of sampling bias on machine learning models
- investigating profit-fairness trade-off in credit scoring with fair machine learning
- performing cost-driven feature selection with multi-objective evolutionary algorithms

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

**10/2014 – 12/2015** — **Research Assistant**, *Institute for Statistical Studies at Higher School of Economics*, Moscow.
- preprocessed survey data for statistical analysis
- conducted the econometric analysis of survey data

## 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.
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: 9.38 of 10 (1.31).
Focused on econometrics and time series. Diploma with honors. 1st place in the 4-year student rating.

## Certificates

**06/2020** — **Udacity Deep Learning Nanodegree:** building CNNs, RNNs and GANs in PyTorch

**05/2020** — **Udacity Machine Learning Engineer Nanodegree:** deploying ML and DL models in AWS

## Technical Stack

| Tools | | |
|---|---|---|
| | **Programming:** Python, R, SQL (interm) | **IDE:** VS Code, Atom, Jupyter, Google Colab |
| | **Deployment:** AWS SageMaker | **MLOps:** Neptune.ai, SageMaker Model Monitor |
| | **Collaboration:** Git, Slack, Trello | **Documents:** LaTeX, Overleaf, MS Office |

| Packages | |
|---|---|
| | **ML:** scikit-learn, pandas, numpy, xgboost, lightgbm, hyperopt, scipy, matplotlib, seaborn |
| | **DL:** pytorch, timm, albumentations, transformers, tensorflow & keras (interm) |

| Algorithms | |
|---|---|
| | **ML:** Boosting (XGB, LGB), Tree-based (RF, DT), Linear (Logreg, OLS), KNN |
| | **DL:** CNNs, RNNs (LSTM, GRU), Transformers (BERT, GPT, vision), Autoencoders |

| Applications | |
|---|---|
| | **Supervised:** classification, regression, time series forecasting |
| | **Unsupervised:** clustering, dimensionality reduction |
| | **CV:** image classification, object detection, image captioning |
| | **NLP:** sentiment analysis, text classification |

## Awards and Achievements

| | |
|---|---|
| 2018 – 2021 | Kaggle Competitions Master (15 medals). Top-1% in Competitions, Notebooks and Discussion |
| 2014 – 2017 | Awarded with Oxford-Russia Fund (2014/15) and E-Fellows.net (2016/17) scholarships |
| 2016 | Prize-winner of the student research paper competition in Computer Science held by HSE |

## Selected ML Competitions

| | |
|---|---|
| 06/2021 | **BMS Molecular Translation**, top-5% (874 teams).<br>Image-to-text translation of chemical molecule structures with CNN-LSTM encoder-decoder models. |
| 02/2021 | **Cassava Leaf Disease Classification**, top-1% (3900 teams).<br>Identifying plant disease types in plant leaf images with CNNs and Vision Transformers. |
| 02/2019 | **Google Analytics Revenue Prediction**, top-2% (3611 teams).<br>Predicting future customer spendings in Google Merchendise Store with gradient boosted trees. |

Code, writeups and further competitions are available at: https://kozodoi.me/kaggle

## Selected Publications

| | |
|---|---|
| 03/2021 | N. Kozodoi, J. Jacob, and S. Lessmann<br>**Fairness in Credit Scoring: Assessment, Implementation and Profit Implications**<br>arXiv:2103.01907 (https://arxiv.org/pdf/2103.01907.pdf) |
| 04/2020 | N. Kozodoi, P. Katsas, S. Lessmann, L. Moreira-Matias and K. Papakonstantinou<br>**Shallow self-learning for reject inference in credit scoring**<br>ECML PKDD 2019 Proceedings, pp. 516-532 (https://doi.org/10.1007/978-3-030-46133-1) |
| 09/2019 | N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis and B. Baesens<br>**A multi-objective approach for profit-driven feature selection in credit scoring**<br>Decision Support Systems, 120 (2019), pp. 106-117 (https://doi.org/10.1016/j.dss.2019.03.011) |

Further papers and full texts are available at: https://kozodoi.me/research

## Software Packages

| | |
|---|---|
| 09/2019 – present | **fairness**: R package for calculating and visualizing fair ML metrics (>12k total downloads) |
| 07/2020 – 02/2021 | **dptools**: Python package with helper functions for data processing and feature engineering |

## Languages

| | | |
|---|---|---|
| English | **Proficient user (C2)** | *IELTS band 8.0 certificate* |
| German | **Advanced user (C1)** | *Language courses at HU Berlin* |
| Russian | **Native speaker** | |

## Skills and Interests

| | | |
|---|---|---|
| Key Skills | ○ Passionate about machine learning and data | ○ Inspired by using ML to improve decision-making |
| | ○ Strong motivation to learn and improve | ○ Excellent presentation skills |
| Hobbies | Football, beach volleyball, piano, scootering | |

Berlin, 11.06.2021