



Nikita Kozodoi

- ✉ n.kozodoi@icloud.com
- ☎ +49 176 *** 35125
- 📍 10627 Berlin, Germany

- 🌐 <https://kozodoi.me> (blog & portfolio)
- 🌐 <https://linkedin.com/in/kozodoi>
- 🌐 <https://github.com/kozodoi>

Education

- 04/2018 – present **PhD in Information Systems**, *Humboldt University*, Berlin.
Research and on machine learning applications for credit risk analytics.
- 10/2015 – 12/2017 **M.Sc. in Economics and Management Science**, *Humboldt University*, Berlin, GPA: 1.30.
Focused on business analytics, data science and applied predictive analytics.
- 09/2014 – 12/2016 **M.Sc. in Economics**, *Higher School of Economics*, Moscow, GPA: 9.58 of 10 (1.21).
Focused on econometrics and time series analysis. Diploma with honors.
- 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 marketing. Diploma with honors.

Work Experience

- 04/2018 – present **Data Science Research Associate**, *Monedo*, Hamburg.
 - developing solutions to mitigate the impact of sampling bias on scoring models
 - performing cost-driven feature selection with multi-objective evolutionary algorithms
 - investigating profit-fairness trade-off in credit scoring with fair machine learning methods
- 07/2016 – 03/2018 **Research Assistant**, *Max Planck Institute for Human Development*, Berlin.
 - collected and prepared a large number of marketing data sets
 - benchmarked managerial heuristics against machine learning algorithms
 - analyzed factors affecting the predictive performance
- 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

Technical Skills

Programming: Python, R, SQL

IDE: VS Code, Atom, Jupyter, Google Colab

Deployment: AWS SageMaker

Monitoring: Neptune, SageMaker Model Monitor

Collaboration: Git, Slack, Trello, Jira

Documents: \LaTeX , MS Office, Apple iWork

Packages	Python: sklearn, pytorch, tensorflow, xgboost, lightgbm, hyperopt, numpy, pandas, matplotlib R: mlr, caret, tidyverse (dplyr, ggplot2), data.table, xgboost, h2o, glmnet, plotly, knitr
Algorithms	ML: GBM (XGB, LGB, Catboost), Tree-based (RF, DT), Linear (Logreg, OLS), KNN DL: CNNs, RNNs (GRU, LSTM), Transformers (BERT), Autoencoders, GANs
Applications	Supervised ML: classification, regression, time series analysis Unsupervised ML: anomaly detection, clustering, dimensionality reduction Computer vision: image classification, object detection, style transfer, image generation Natural language processing: sentiment analysis, text generation

Awards and Achievements

- 2018 – 2021 Kaggle Competitions Master (14 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

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

Software Packages

- 07/2020 – present **dptools**: Python package with helper functions for data processing and feature engineering
- 09/2019 – present **fairness**: R package for computing and visualizing metrics of algorithmic fairness

Selected ML Competitions

- 03/2021 **RANZCR Catheter and Line Position Challenge**: top-5% (solo, 1547 teams)
- 02/2021 **Cassava Leaf Disease Classification**: top-1% (2-people team, 3900 teams)
- 08/2020 **SIIM-ISIC Melanoma Classification**: top-1% (5-people team, 3314 teams)
- 10/2019 **IEEE-CIS Fraud Detection**: top-3% (2-people team, 6381 teams)
- 02/2019 **Google Analytics Revenue Prediction**: top-2% (2-people team, 3611 teams)
- 12/2018 **PLAsTiCC Astronomical Classification**: top-5% (3-people team, 1094 teams)
- 08/2018 **Home Credit Default Risk Prediction**: top-4% (solo, 7198 teams)

Selected Publications

- 04/2020 N. Kozodoi, P. Katsas, S. Lessmann, L. Moreira-Matias and K. Papakonstantinou
Shallow self-learning for reject inference in credit scoring
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
A multi-objective approach for profit-driven feature selection in credit scoring
Decision Support Systems, 120 (2019), pp. 106-117 (<https://doi.org/10.1016/j.dss.2019.03.011>)

Languages

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

Skills and Interests

- | | | |
|------------|--|---------------------------------|
| Key Skills | ◦ Passionate about machine learning and data | ◦ Excellent presentation skills |
| | ◦ Strong motivation to learn and improve | ◦ Good at meeting deadlines |
| Hobbies | Beach volleyball, piano, scootering | |