



# Nikita Kozodoi

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

## Education

- 04/2018 – present **PhD in Information Systems**, *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.

## Work 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

## Technical Skills

**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	<b>ML:</b> scikit-learn, pandas, numpy, xgboost, lightgbm, hyperopt, scipy, matplotlib, seaborn <b>DL:</b> pytorch, timm, albumentations, transformers, tensorflow & keras (interm)
Algorithms	<b>ML:</b> Boosting (XGB, LGB), Tree-based (RF, DT), Linear (Logreg, OLS), KNN <b>DL:</b> CNNs, RNNs (LSTM, GRU), Transformers (BERT, GPT, vision), Autoencoders
Applications	<b>Supervised:</b> classification, regression, time series forecasting <b>Unsupervised:</b> clustering, dimensionality reduction <b>CV:</b> image classification, object detection, image captioning <b>NLP:</b> sentiment analysis, text classification

## 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

---

## 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

- Computer vision **Cassava Leaf Disease Classification:** top-1% (3900 teams)
- Computer vision **SIIM-ISIC Melanoma Classification:** top-1% (3314 teams)
- Tabular data **Google Analytics Revenue Prediction:** top-2% (3611 teams)
- Tabular data **IEEE-CIS Fraud Detection:** top-3% (6381 teams)
- Time series **PLAsTiCC Astronomical Classification:** top-5% (1094 teams)
- Solution writeups and code are available at <https://kozodoi.me/kaggle>

---

## Selected Publications

- 03/2021 N. Kozodoi, J. Jacob, and S. Lessmann  
**Fairness in Credit Scoring: Assessment, Implementation and Profit Implications**  
arXiv:2103.01907 (<https://arxiv.org/pdf/2103.01907.pdf>)
- 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>)

---

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

---

## 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 |   |