



# 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 machine learning and deep 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 / 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 methods
  - performing cost-driven feature selection with multi-objective evolutionary algorithms
  - teaching machine learning and supervising M.Sc. dissertations on research-related topics
- 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:**  $\text{\LaTeX}$ , MS Office, Apple iWork

Packages	<b>Python:</b> sklearn, pytorch, tensorflow, xgboost, lightgbm, hyperopt, numpy, pandas, matplotlib <b>R:</b> mlr, caret, tidyverse (dplyr, ggplot2), data.table, xgboost, h2o, glmnet, plotly, knitr
Algorithms	<b>ML:</b> Boosting (XGB, LGB, CatBoost), Tree-based (RF, DT), Linear (Logreg, OLS), KNN <b>DL:</b> CNNs, RNNs (LSTM, GRU), Transformers (BERT variants), Autoencoders
Applications	<b>Supervised ML:</b> classification, regression, time series forecasting <b>Unsupervised ML:</b> clustering, anomaly detection, dimensionality reduction <b>Computer vision:</b> image classification, object detection, image captioning <b>Natural language processing:</b> 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

- 09/2019 – present **fairness:** R package for calculating and visualizing fair ML metrics (>11k total downloads)
- 07/2020 – 02/2021 **dptools:** Python package with helper functions for data processing and feature engineering

---

## 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>)

---

## 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
  - Excellent presentation skills
  - Strong motivation to learn and improve
  - Good at meeting deadlines
- Hobbies
- Football, beach volleyball, piano, scootering