Konstantin Burkin

konstantin.burkin.m@gmail.com | kburkin | kozersky | konstantinburkin.github.io

SUMMARY: Data Scientist in the largest online retail business. 3+ years of production experience.
Computer Science PhD and alumnus of Lomonosov Moscow State University.

WORK EXPERIENCE

- **Senior Data Scientist** *June 2024 – Present* • **ecom.tech**
  - Project: Courier shifts optimization
  - Stack: Python, Polars, DuckDB, CatBoost, Pyomo
    - Calibrated courier shifts with money-loss elastic model and linear optimization with constraints.
    - Resulting approach deployed to production, reducing monthly losses by millions ₽.

  - Project: Forecasting demand for parcel shipping
  - Stack: Polars, DuckDB, SQL, S3, Greenplum, GitLab, Poetry
    - Deployed modules on server with full data engineering pipeline, model fitting and prediction.
    - Automated model retraining and inference with scheduled runs reducing employee labor time by 5%.

- **Middle Data Scientist** *Aug 2022 – June 2024* • **McDonalds**
  - Project: Analysis of customer reviews with NLP
  - Stack: Python, ruBERT, Yandex-GPT
    - Automated suggestions for app improvement based on scrapped reviews from AppStore and PlayMarket.
    - Preselected informative reviews for each function of app and aggregated them using Yandex-GPT.

  - Project: Sales forecasting
  - Stack: Python, PyTorch, SQL, CatBoost, SARIMA, Airflow, MLflow, Git, DVC
    - Developed LSTM for univariate time series prediction to decrease retraining costs and improved WAPE by 6%.
    - Modeled products similarity via graph architecture to predict sales of new products using GNN.
    - Developed baseline model for predictions of unpopular products. Improved MAE by 4%.
    - Engineered features for boosting models to improve predictions of regular and promo sales.

  - Project: Mentoring
    - Mentored intern for 4 months until his promotion to Junior Data Scientist position.

- **Junior Researcher / Data Scientist** *Sep 2021 – Aug 2022* • **Webiomed**
  - Project: Risk prediction for cardiovascular patients
  - Stack: Python, Git, Bash, scikit-learn, imblearn, Optuna, MLxtend, CatBoost, Pandas, NumPy
    - Conducted model interpretation, feature selection and importance analysis to ensure medical validity.
    - Increased Recall by 9% by tuning models with respect to F_2 metric.

EDUCATION

- **PhD program in Machine Learning** *Nov 2023 – Present* • **Higher School of Economics University**
  - Project: Parsing medical text records using NLP
    - Parsing medical records to extract textual description of patient features for NLP analysis.
    - Fine-tuned BERT-based encoder to classify patients with risk of disease progression.
    - Uplift modeling of medicine intake for patients with genetic mutations

- **Published research** scholar.google.com

- **BSc & MSc in Fundamental and Applied Chemistry** *Sep 2017 – Aug 2023* • **Lomonosov Moscow State University**
  - GPA: 4.97/5, Red diploma
  - Academic council Scholarship: top-10 MSU students for scientific achievements