


# Konstantin Burkin

✉ [konstantin.burkin.m@gmail.com](mailto:konstantin.burkin.m@gmail.com) | [in kburkin](https://www.linkedin.com/in/kburkin) | [kozersky](https://kozersky.com) | [konstantinburkin.github.io](https://konstantinburkin.github.io)

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

## WORK EXPERIENCE

---

- **Middle Data Scientist/MLOps Engineer** 2024 – Present • [samokat.tech](https://samokat.tech)
  - Project: Forecasting demand for parcel shipping
  - Stack: Python, Polars, DuckDB, SQL, GitLab
    - Developed Boosting prediction models and data cleaning technique, improving MAPE by 10%.
    - Deployed project repository on server with scheduled runs and engineered data flow to data storage.
- **Middle Data Scientist** 2023 – 2024 • [McDonalds.tech](https://McDonalds.tech)
  - 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 eliminate scheduled training and improve 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.
- **Junior Machine Learning Researcher** 2021 – 2023 • [Higher School of Economics University](https://HigherSchoolofEconomicsUniversity.com)
  - Project: Risk prediction for cardiovascular patients
  - Stack: Python, Git, Bash, scikit-learn, imblearn, Optuna, MLxtend, CatBoost, Pandas, NumPy
    - Demonstrated biomarkers' predictive capabilities (up to 5% AUROC increase).
    - Increased Recall by 9% by tuning models with respect to  $F_2$  metric.
    - Published paper at peer-reviewed journal. 


## EDUCATION

---

- **PhD program in Machine Learning** 2023 – Present • [Higher School of Economics University](https://HigherSchoolofEconomicsUniversity.com)
  - Project: Parsing unstructured medical data
- **Program: Neural networks in research** 2022 – 2023 • [Lomonosov Moscow State University](https://LomonosovMoscowStateUniversity.com)
  - Scholarship: top-25 based on ML competition and academic results
- **BSc & MSc in Fundamental and Applied Chemistry** 2017 – 2023 • [Lomonosov Moscow State University](https://LomonosovMoscowStateUniversity.com)
  - GPA: 4.97/5, Red diploma
  - Academic council Scholarship: top-10 MSU students for scientific achievements

## INDIVIDUAL PROJECTS

---

- **Delivery Club Sales Prediction** 
  - Stack: Python, scikit-learn, CatBoost, Pandas, PyTorch, Pandas, NumPy, Plotly
    - Compared effectiveness CatBoost and LGBM models for weekly sales prognosis.
    - Compared models for univariate time series. SARIMA outperformed LSTM in accuracy and training time.