konstantin.burkin.m@gmail.com kburkin kozersky konstantinburkin.github.io +7(900)-150-7887

Konstantin Burkin

SUMMARY

- Data Scientist | E-commerce - Retail - Logistics
- 3+ years of production experience
- PhD in Computer Science

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.
    - Conducted A/B test measuring ARPPU and LTV, which reduced operational expenses by millions $ annually.

  - Project: Supply chain forecasting
  - Stack: Polars, DuckDB, SQL, S3, Greenplum, Airflow, GitLab, CatBoost, Poetry
    - Improved prognosis MAPE by 8% and simplified model monitoring by replacing autoregression with boosting.
    - Deployed and automated model retraining and inference reducing employee labor time by 5%.

- **Middle Data Scientist** *Aug 2022 – June 2024* • **McDonalds**
  - Project: Customer reviews analysis for ecom application
  - Stack: Python, BERT, Transformer, Hugging Face
    - Generated suggestions for app improvement based on AppStore reviews using pretrained Transformer model.
    - Monitored users' satisfaction level by categorizing reviews and analyzed sentiment with fine-tuned BERT model.

  - 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 in healthcare
  - Stack: Python, Git, Bash, scikit-learn, imblearn, Optuna, MLxtend, CatBoost, Pandas, NumPy
    - Improved high-risk patients' detection: increased Recall by 9% by tuning models with respect to F_2 metric.
    - Conducted model interpretation, feature selection and importance analysis to ensure medical validity.

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 in Fundamental Chemistry & MSc in Biochemistry** *Sep 2017 – Aug 2023* • **Lomonosov Moscow State University**
  - GPA: 4.97/5, scholarship for scientific achievements