

# GRIGORIY OVCHINNIKOV

Data Scientist

Georgia, Tbilisi

✉ [ogoj@protonmail.com](mailto:ogoj@protonmail.com)

in [grigoriy-o-a679501b5/](https://www.linkedin.com/in/grigoriy-o-a679501b5/)

[Personal site](#)

🔗 [JointEntropy](#)

Data Scientist with more than 5 years of data analysis and machine learning experience. Strong knowledge of classical, advanced ML algorithms, applying deep learning and statistics especially in AdTech domain. Proficiency in working with ML frameworks and databases for business problem-solving.

## TECHNICAL SKILLS

---

**Languages:** Python, GoLang, SQL

**Machine learning:** Scikit-learn, XGBoost, LightGBM, CatBoost, PyTorch, SciPy, NumPy, Pandas, Matplotlib, Plotly, Seaborn, pulearn, shap, MLflow

**Databases & Message broker systems & Workflow orchestration:** ClickHouse, PostgreSQL, MongoDB, Cassandra, HBase, Apache Kafka, Redis, Airflow

**Version Control & Deploy & Infrastructure:** Git, GitLab CI/CD, Docker(compose, swarm)

**GCP:** Cloud Functions, Cloud Run, Cloud Storage, PubSub, VertexAI

**Big data:** Hadoop, Hive, Pyspark

**Natural Language Processing:** gensim, nltk, transformers, pytorch-lightning

**Web scraping:** Scrapy, Puppeteer, lxml

**Project management:** Jira, Confluence, Trello

**Other:** EDA, Data cleaning, Visualization, Modelling, Propensity Modelling, Model interpretation, Natural Language Processing, Statistics, Experimental design, Hypothesis testing, MVP

## EXPERIENCE

---

### Go Net | DSP

February 2023 – Present

*Senior ML Engineer*

*Limmasol, Cyprus*

- Implemented **core machine learning based mechanics for DSP**: max CPC and target CTR bidding strategies based on control theory methods. Designed and implemented a **high-throughput scalable ML model inference service using Go**.
- Designed system for contextual targeting feature based on natural language processing and large language modelling(with transformers and PyTorch Lightning).
- Led technical interviews for Machine Learning Engineer roles, enhancing team capabilities.
- MLOps operations: model monitoring, quality assessment, conducting A/B testing. Providing data analysis of collected data to guide decision-making processes.
- Pipeline stack**: Python, GoLang, ClickHouse, Kafka, Aerospike, transformers, PyTorch, Redis, Catboost, vowpal-wabbit, gRPC, Prometheus

### Platforma

November 2020 – December 2022

*Data Scientist | ML Engineer*

*Moscow, Russia*

- Implemented state-of-the-art methods for a look-alike modeling task with positive unlabeled (PU) learning using PySpark and PyTorch with a **significant uplift in offline and online metrics**.
- Built an efficient, accurate and scalable ML solution for labeling clickstream data which **increased accuracy and lower inference time on telecom data in more than hundreds of million devices** using PySpark and Hadoop.
- Suggested and implemented an NLP-inspired(item2vec) approach for website representation learning based on clickstream data, giving several advantages over the previous approach.
- Pipeline stack**: Python, Hadoop, Pyspark, Airflow, PyTorch.

### S7 TechLab

October 2018 – October 2020

*Junior Data Scientist*

*Moscow, Russia*

- Conducted research and data analysis to address predictive maintenance task. Provided estimates of remaining useful life through supervised machine learning in the big data department.
- Developed ML models for improving accuracy of decision support system using cluster analysis and supervised multiclass classification.
- Scraping external data with Python, Building MVP with R Shiny, Python Flask-restx.
- Pipeline stack**: Python, Airflow, PySpark, S3, PostgreSQL, MongoDB, Cassandra, Kafka.

## EDUCATION

---

### National Research Nuclear University (MEPhI)

*Master's degree in Applied Mathematics and Informatics(incomplete)*

**September 2018 – June 2020**

*Moscow, Russia*

### Volgograd State Technical University(VSTU)

*Bachelor's degree in Informatics and Computer Engineering*

**September 2014 – June 2018**

*Volgograd, Russia*

## CERTIFICATIONS

---

- Google Cloud, Big Data and Machine Learning Fundamentals [November 2022]
- TensorFlow on Google Cloud [November 2022]
- Drawing conclusions from data, MIPT, Yandex [October 2020]

## Languages

---

- English - B2(Upper intermediate)
- Russian – Native