

# GRIGORIY OVCHINNIKOV

Data Scientist | ML Engineer

Georgia, Tbilisi

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

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

[Personal site](#)

🔗 [JointEntropy](#)

Data Scientist and ML Engineer with 4 years of data analysis and machine learning experience. Strong knowledge of classical, advanced ML algorithms, applying deep learning and statistics. Proficiency in working with ML frameworks and databases for business problem-solving.

## TECHNICAL SKILLS

---

**Languages:** Python, SQL

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

**Databases:** PostgreSQL, MongoDB, Cassandra, HBase

**Deployment:** Git

**Google Cloud Platform:** Cloud Functions, Cloud Run, Kubernetes Engine

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

**Natural Language Processing:** gensim, nltk, transformers

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

**Infrastructure:** Docker

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

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

## EXPERIENCE

---

### Platforma

November 2020 – Present

*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.

## SIDE PROJECTS

---

**CatchBlogger** [🔗](#) | Co-founder and developer of YouTube influence marketing platform.

06 2022

- Responsible for whole data management pipeline: extraction, transformation and loading.
- Pipeline stack:** Python, Puppeteer, GCP: Cloud Functions, Cloud Scheduler, Cloud Storage

## EDUCATION

---

**National Research Nuclear University (MEPhI)**

September 2018 – June 2020

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

*Moscow, Russia*

**Volgograd State Technical University(VSTU)**

September 2014 – June 2018

*Bachelor's degree in Informatics and Computer Engineering*

*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