

# ГРИГОРИЙ ОВЧИННИКОВ

Data Scientist | ML Engineer

Москва, Россия

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

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

[Personal site](#)

🔗 [JointEntropy](#)

Data Scientist и ML Engineer с более чем 4 годами опыта анализа данных, машинного обучения. Уверенные знания классических и продвинутых алгоритмов машинного обучения, применения глубокого обучения и статистики. Богатый опыт работы с различными ML фреймворками и разработки полного ML пайплайна для решения бизнес-задач.

## ТЕХНИЧЕСКИЕ СКИЛЛЫ

---

**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:** Scrappy, 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

## ОПЫТ РАБОТЫ

---

### Платформа больших данных

Ноябрь 2020 – по н.в.

*Data Scientist | ML Engineer*

*Москва, Россия*

- Реализовывал state-of-the-art методы решения задачи look-alike моделирования с помощью positive unlabeled (PU) обучения используя PySpark и PyTorch, что дало **существенный uplift на оффлайн и онлайн метриках**.
- Разработал эффективное, точное и масштабируемое ML решения для разметки данных clickstream трафика, которое **увеличило точность и снизило время прогнозирования на данных телеком трафика с сотнями миллионами устройств** используя PySpark и Hadoop.
- Предложил и реализовал вдохновленный NLP методами подход для обучения эмбедингов веб-сайтов на основе данных clickstream трафика, который существенно превзошёл предыдущее решение.
- **Стэк:** Python, Hadoop, Pyspark, Airflow, PyTorch.

### S7 TechLab

Октябрь 2018 – Октябрь 2020

*Junior Data Scientist*

*Москва, Россия*

- Проводил исследование и анализировал данные для решения задачи predictive maintenance. Участвовал в построение ML моделей для оценки Remaining Useful Life(RUL) с помощью обучения с учителем.
- Разрабатывал ML модели для повышения точности систем помощи принятия решения, применяя кластерный анализ и supervised мультиклассовую классификацию.
- Собирал данных из внешних источников с помощью Python, и участвовал в построении MVP с помощью фреймворка R Shiny, Python Flask-restx.
- **Стэк:** Python, Airflow, PySpark, S3, PostgreSQL, MongoDB, Cassandra, Kafka.

## ПЕТ-ПРОЕКТЫ

---

### CatchBlogger | Сооснователь и разработчик платформы для поиска инфлюенсеров на YouTube **06 2022**

- Отвечал за весь ETL пайплайн обработки данных
- **Стэк:** Python, Puppeteer, GCP: Cloud Functions, Cloud Scheduler, Cloud Storage

## ОБРАЗОВАНИЕ

---

**Национальный исследовательский ядерный университет**

**Сентябрь 2018 – Июнь 2020**

*Степень магистра, прикладная математика и информатика(незавершённая)*

*Москва, Россия*

**Волгоградский государственный технический университет**

**Сентябрь 2014 – Июнь 2018**

*Степень бакалавра информатики и вычислительной техники*

*Волгоград, Россия*

## СЕРТИФИКАТЫ

---

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

## ЯЗЫКИ

---

- Английский язык - B2(Upper intermediate)
- Русский язык – родной