Grigoriy Ovchinnikov

Georgia, Tbilisi

✉ ogoje@protonmail.com   🔗 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 and statistics. Proficiency in working with ML frameworks and databases for business problem-solving.

TECHNICAL SKILLS

Languages: Python, SQL, Bash
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
Middle Data Scientist/ML Engineer   *Moscow, Russia*

- Department of data analysis and modeling.
- Look-alike modelling, positive unlabeled(PU) learning.
- User profiling using clickstream data analysis.
- Developing approaches for building user interests vectors.
- Adapting item2vec approach for host representation learning.
- Development of an efficient, accurate and scalable ML solution for clickstream classification on PySpark
- **Pipeline stack**: Python, Hadoop, Pyspark, Airflow, PyTorch.

S7 TechLab 🔗   October 2018 – October 2020
Junior Data Scientist   *Moscow, Russia*

- Big data department.
- Predictive maintenance task, remaining useful life estimation.
- Development of decision support system with 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, 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]
- Drawing conclusions from data, MIPT, Yandex [October 2020]