DMITRY PROTASOV

telegram: @vosatorp
@ dmitry.protasov@gmail.com
github

ABOUT ME

Data Science Engineer with 2+ years of experience in Data Science. Expertise in end-to-end ML workflows: problem scoping, data preparation, model development, and deployment. Strong academic background (MIPT, Yandex School of Data Analysis) and competitive programming experience.

EDUCATION

Bachelor's and Master's in Computer Science

Moscow Institute of Physics and Technology 2018–2022, 2023–2025

- Relevant courses: ML, Algorithms, Python, C++, Databases, Mathematics
- Yandex School of Data Analysis (YSDA):
  - Generative Models: Gained knowledge in GANs, Normalizing Flows, Diffusion Models, VQ-VAE.
  - Computer Vision, NLP, RL, DL, Speech Processing, Self-driving Cars, Efficient DL.

EXPERIENCE

Data Science Engineer

Avito October 2023 - Present

- Boosted the classification macro-precision of the image-based vehicle classifier from 55% to 87% by leveraging an EfficientNet/ViT backbone.
- Built a vehicle fraud detection system, automating over 500 rejections daily with 98% accuracy.
- Developed a Metric Learning-based vision model combined with CatBoost for comparing cars in photos and detecting fake listings, achieving 30% recall on a specialized dataset.
- Enhanced the vin2param model: automated PostgreSQL updates, dataset refresh, and model retraining via S3; expanded functionality to specialized machinery and added Grafana monitoring.
- Developed a segmentation model for detecting vehicle damages (scratches, rust, putty marks).

Machine Learning Researcher/Engineer

AIRI November 2022 - October 2023

- Built ML models for quantum chemistry tasks using reinforcement learning (SAC, TQC, PPO) and Graph Neural Networks (GNNs).
- Improved optimization quality by 15% for molecular trajectory processing.
- Contributed to the open-source schnetpack library and refactored code to enhance performance.

Data Analyst

Yandex September 2021 - December 2021

- Analytics in Yandex Search Proficit Team. Stack: YQL (SQL), Clickhouse, MapReduce

TECHNOLOGIES, SKILLS AND OTHER

- **Programming Languages**: Python, C++, SQL.
- **ML Frameworks**: PyTorch, Scikit-learn, HuggingFace, Numpy, Pandas, Torchaudio, Librosa.
- **Computer Vision**: OpenCV, Torchvision, Ultralytics
- **MLOps and Cloud Solutions**: Docker, Kubernetes, CI/CD pipelines, AWS (EC2, S3), Grafana.
- **Databases and Backend Tools**: PostgreSQL, Vertica, REST API development.
- **Algorithms and Data Structures**

COMPETITIVE PROGRAMMING

- **Codeforces:** 2150 (highest rating)
- **Atcoder:** 1950 (highest rating)
- **ICPC Semifinalist:**
  - 2019 - 1st degree diploma
  - 2021 - 3rd degree diploma
- **Winner of Russian Olympiad in Informatics, 2018**

PROJECTS

- Audio2MIDIBot: Developed a Telegram bot for transcribing music into notes (200+ monthly users)
- Dice Detection and Segmentation Bots: Created Telegram bots for detecting, classifying, and segmenting dice using Ultralytics YOLO models. Applied advanced synthetic data generation techniques to enhance performance and address data scarcity. [DICE Bot], [Triangle Dice Bot], [Digits Bot], [GitHub].
- Developed a tg bot for detecting acne using YOLO-based models. [Acne Detector Bot]
- Participated in the LCT Fest 2024 Hackathon, developing solutions for Named Entity Recognition (NER) extraction from entities.
- Forecasting FMRI (brain scans) with 3D-transformers, analyzed video-like sequences.

PUBLICATIONS

- Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials (arxiv.org/abs/2406.14347)
- Coverings of planar and three-dimensional sets (arxiv.org/abs/2210.12394);